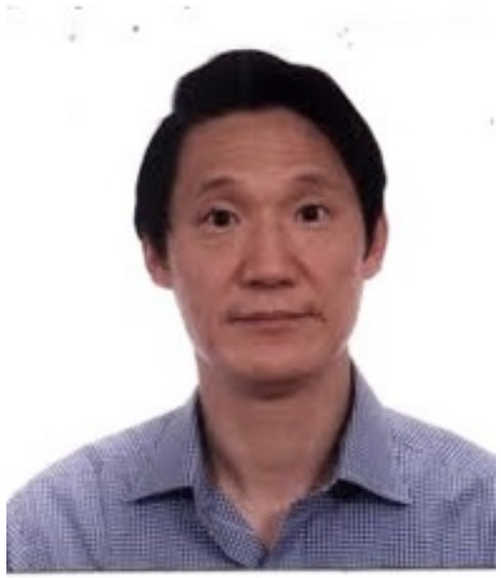
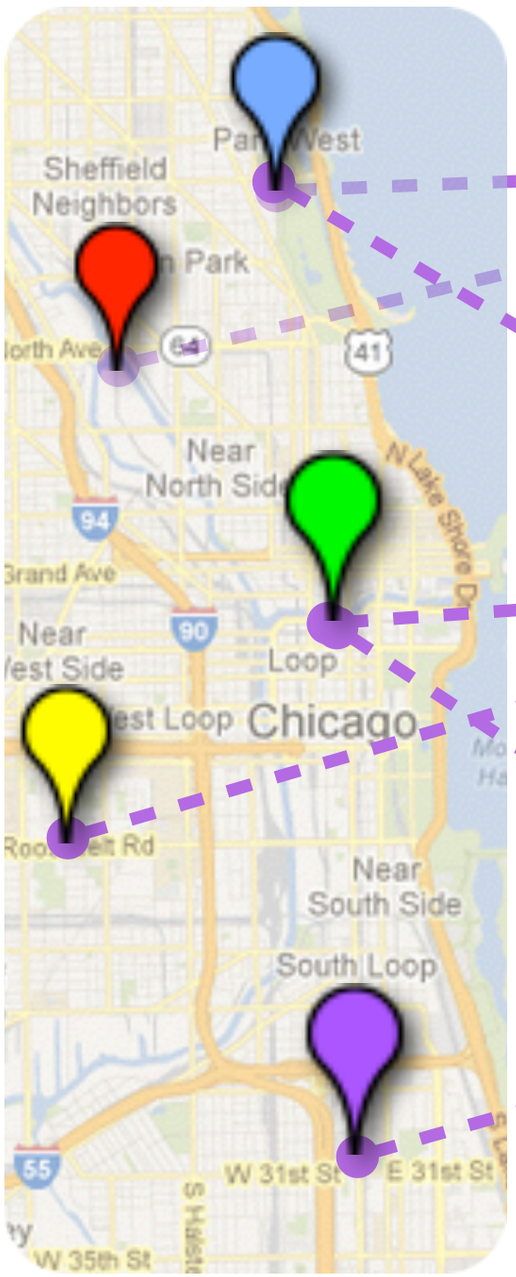
Predicting Social Links for New Users across Aligned Heterogeneous Social Networks



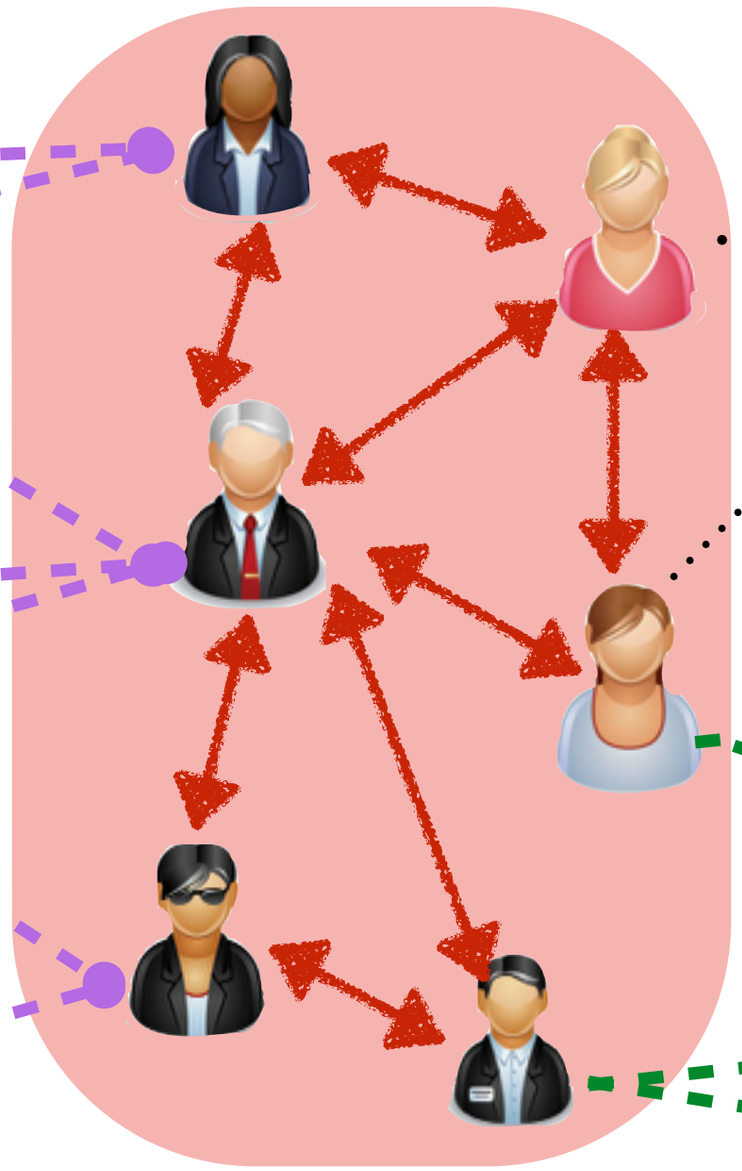
Jiawei Zhang Xiangnan Kong Philip S. Yu

University of Illinois at Chicago

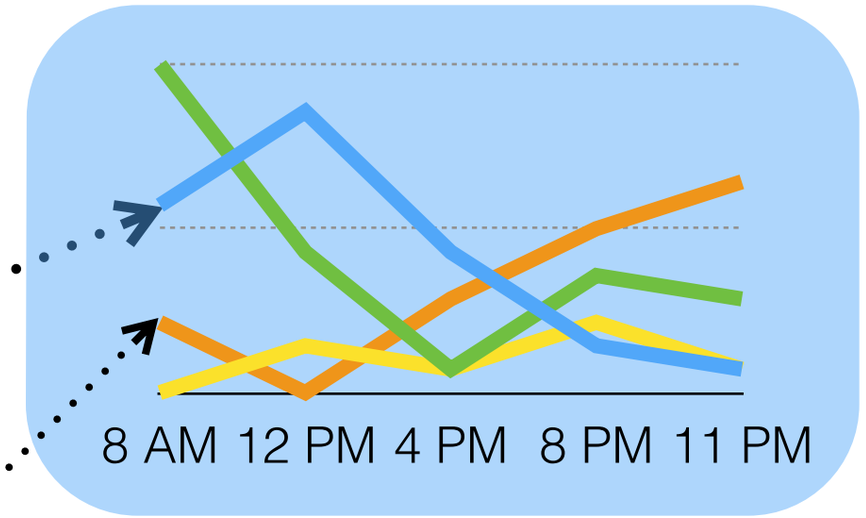
Locations



Social Links



Temporal Activities



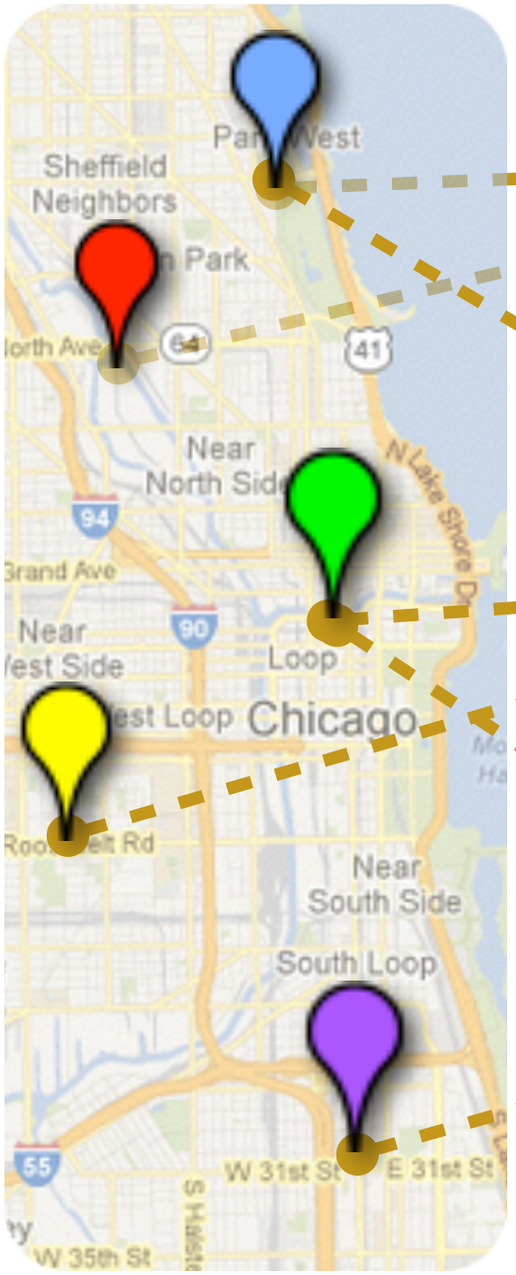
Contents: Tweets

Social Network:

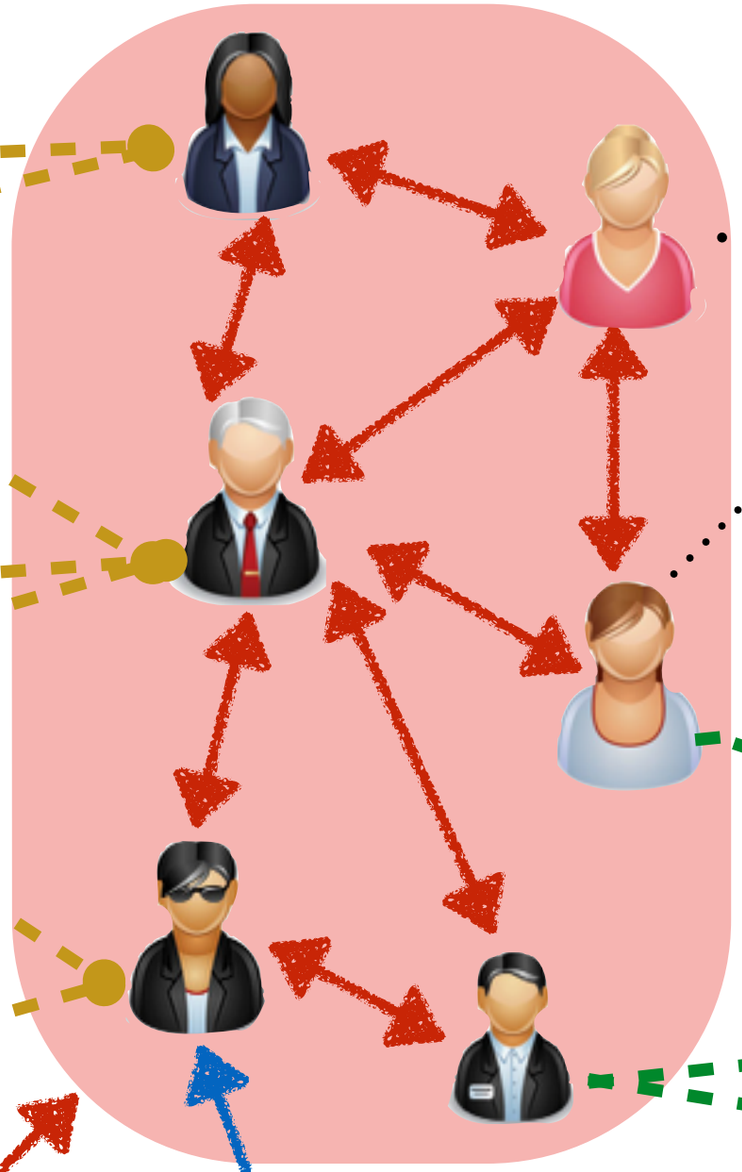
Who Where What When

Problem Description

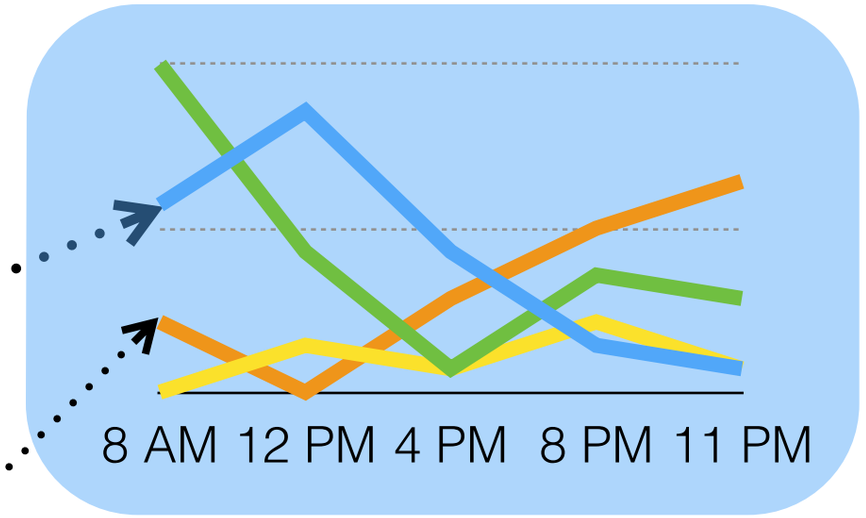
Locations



Social Links



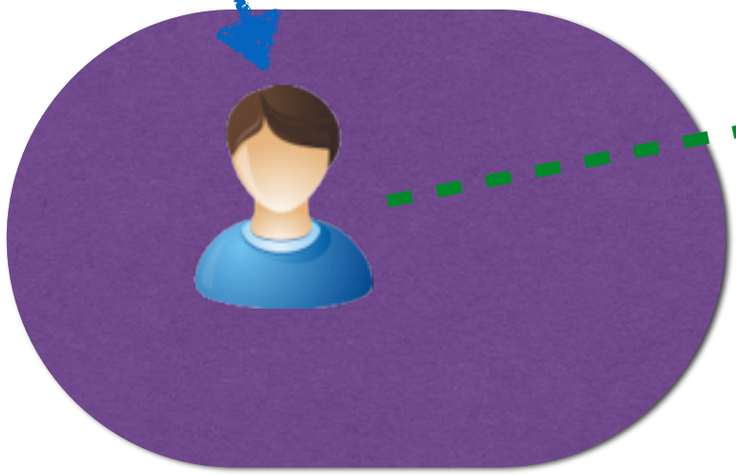
Temporal Activities



Contents: Tweets

old user

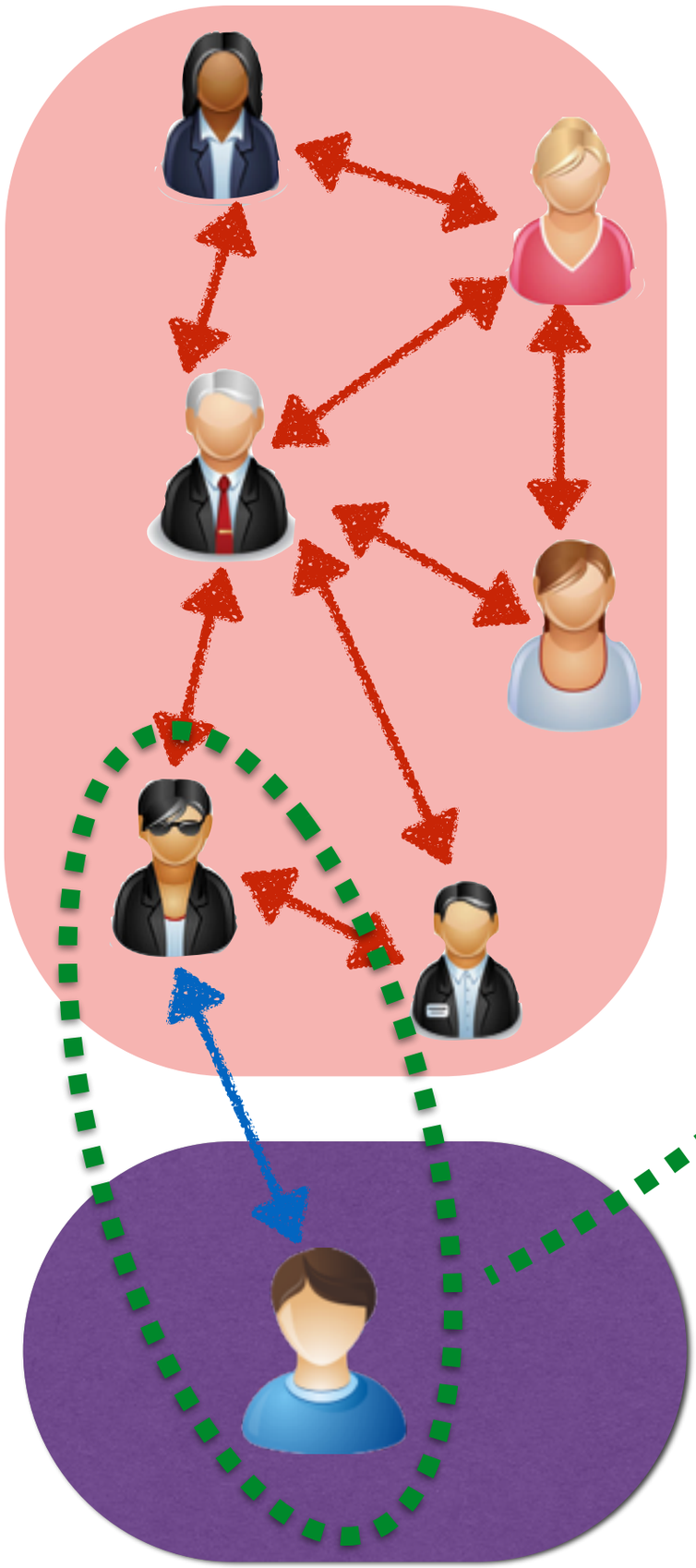
new user



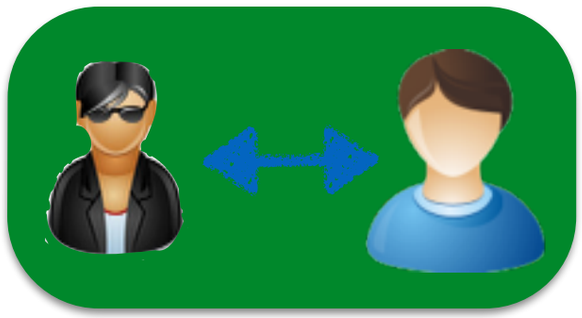
Solve Challenge 1:

Lack of Training Instances

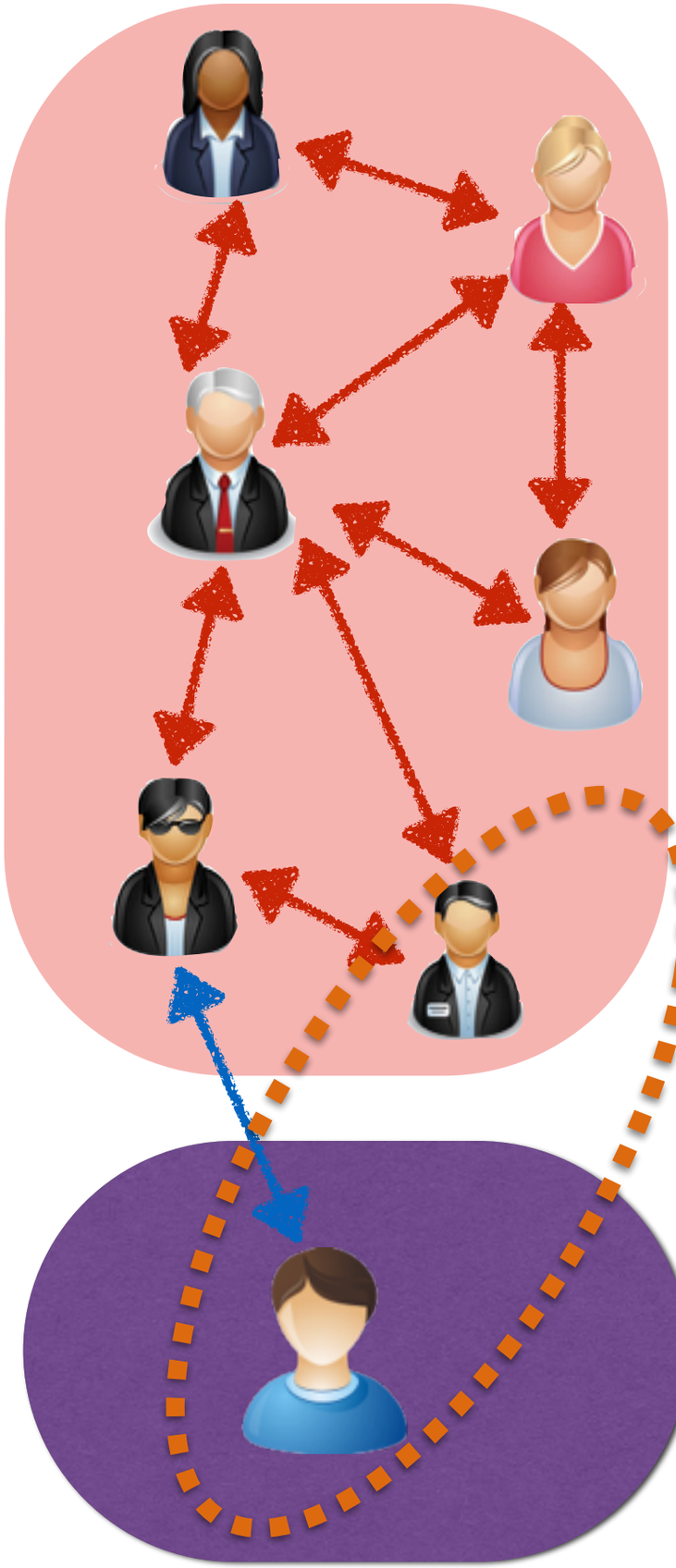
Social Links



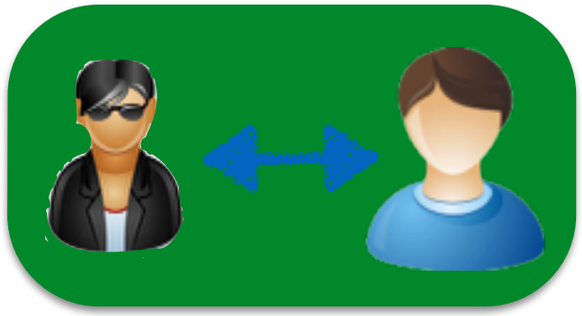
training set



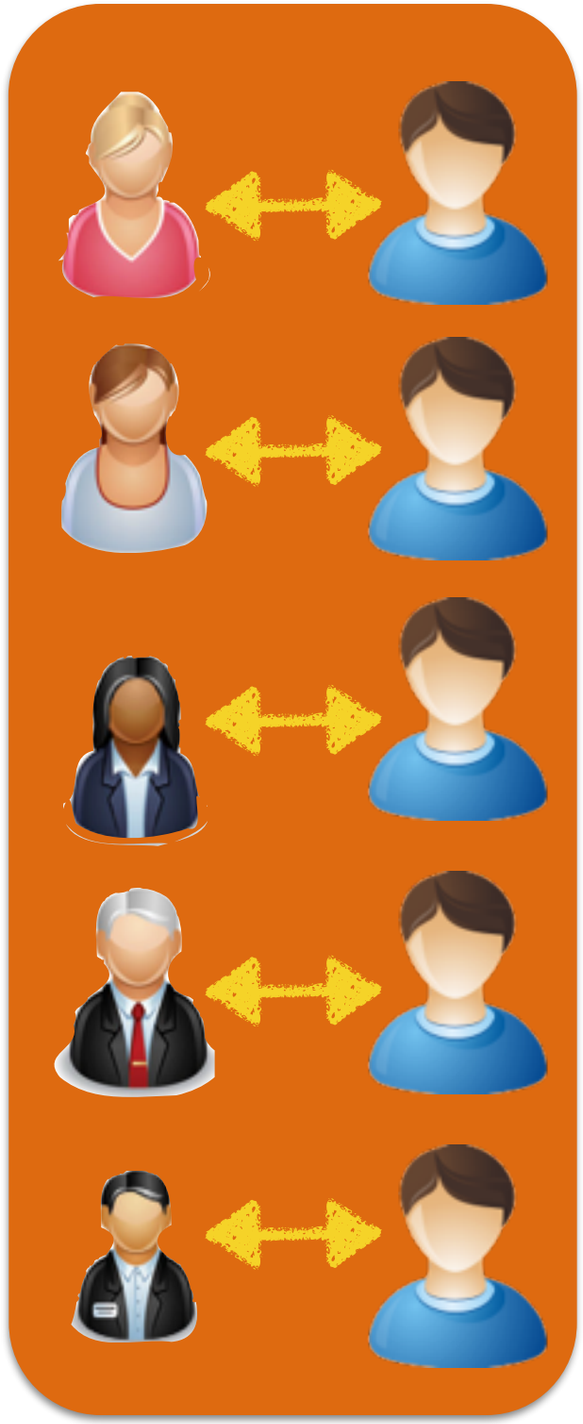
Social Links



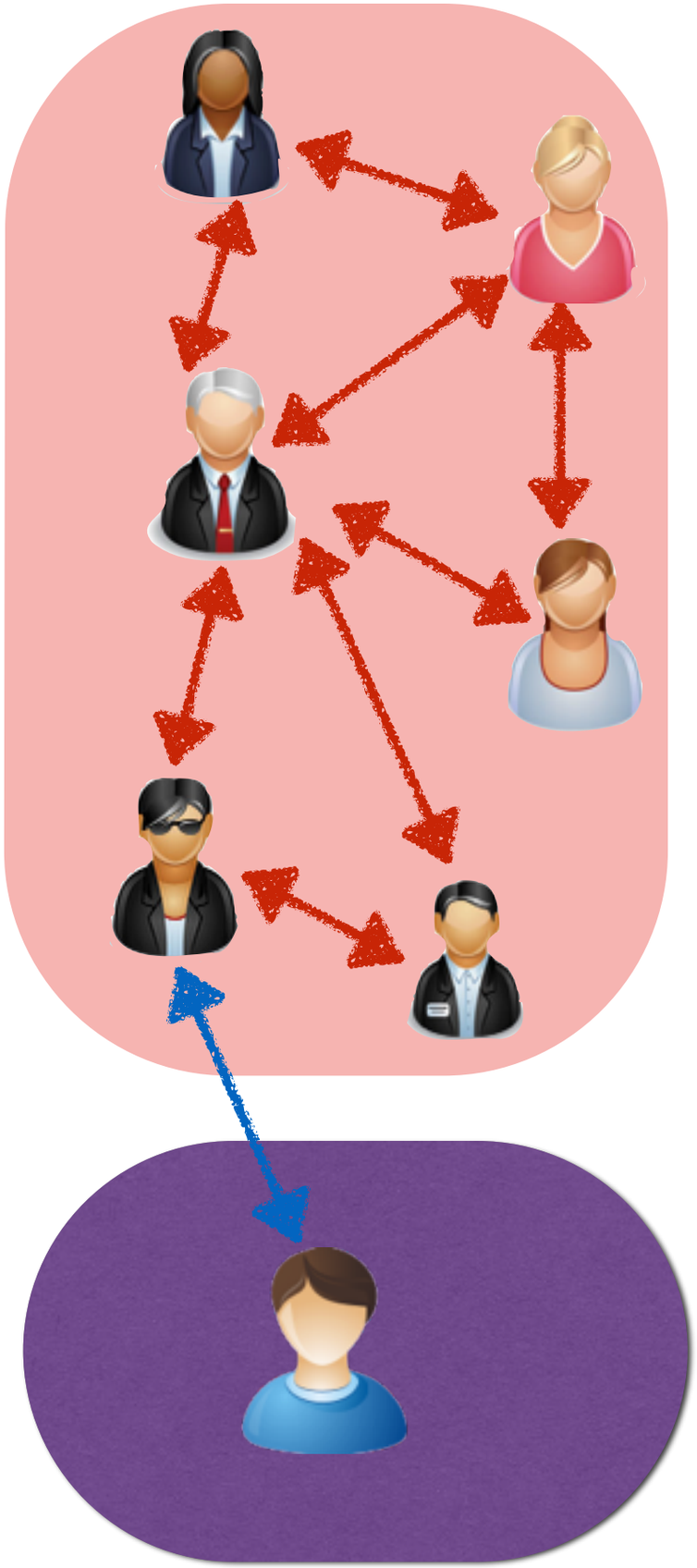
training set



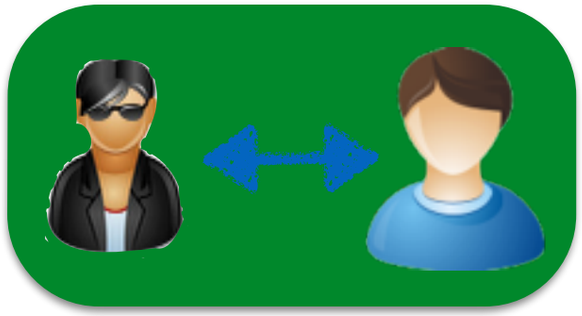
test set



Social Links



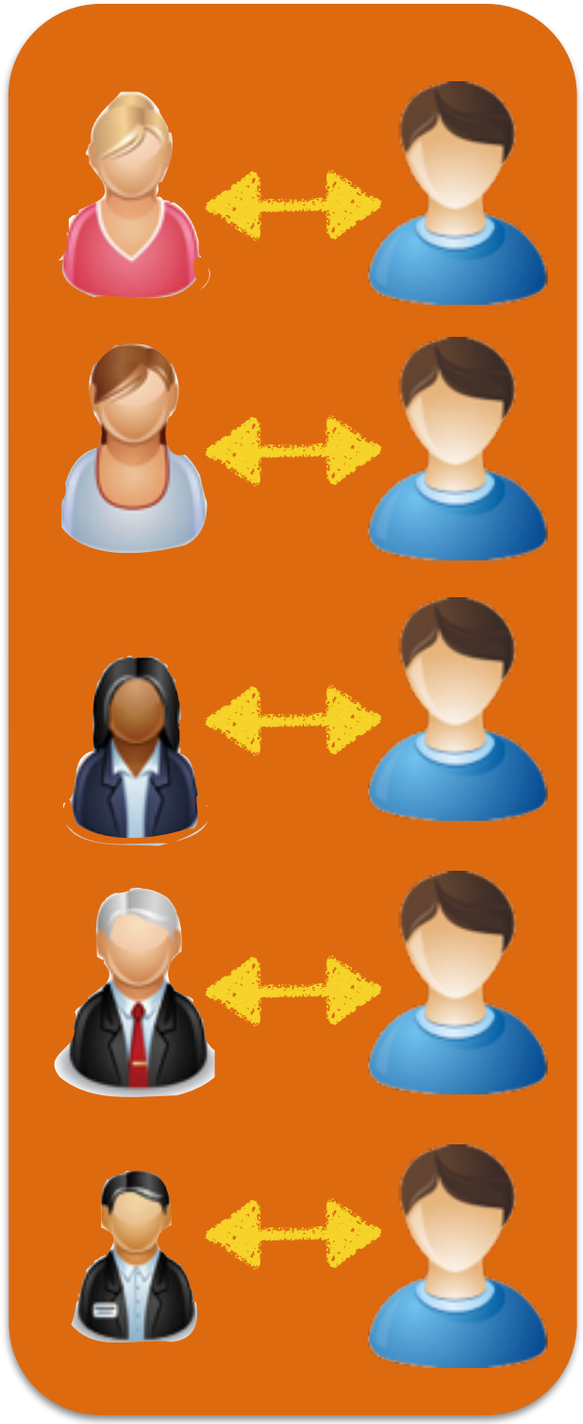
training set



small training set

large test set !!!

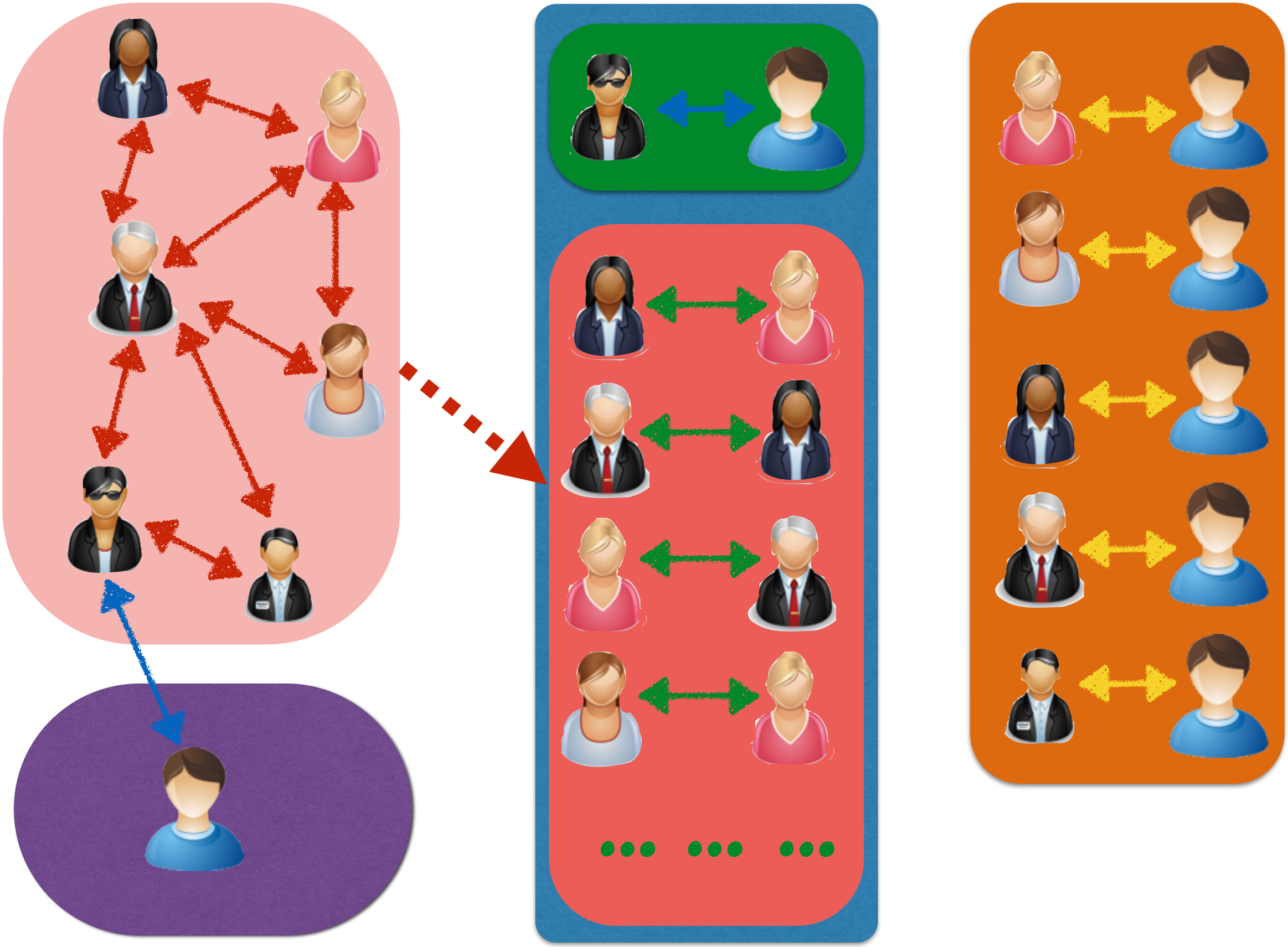
test set



Social Links

training set

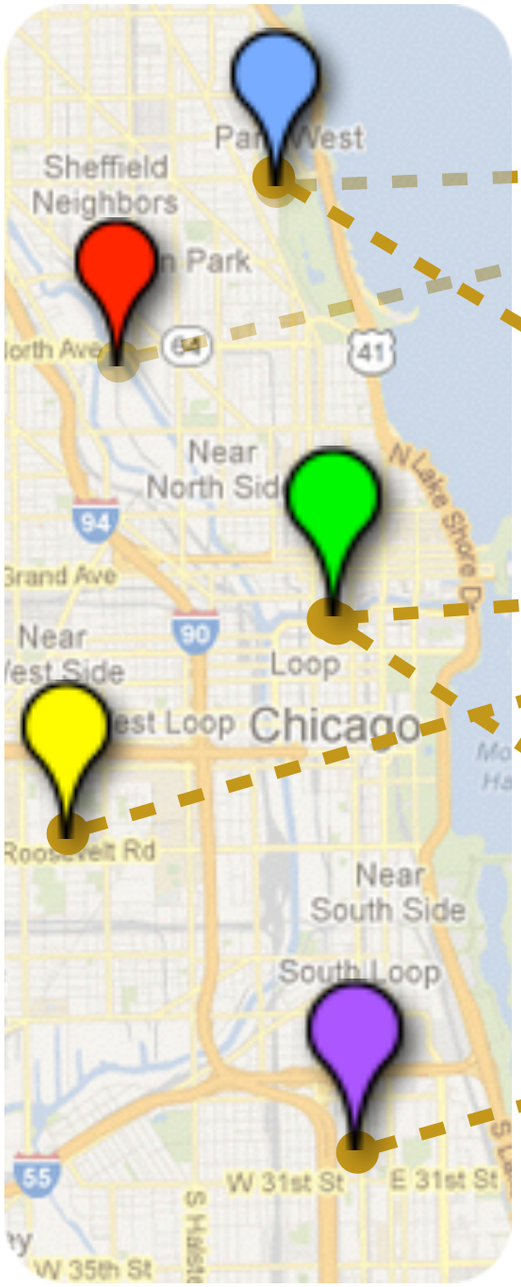
test set



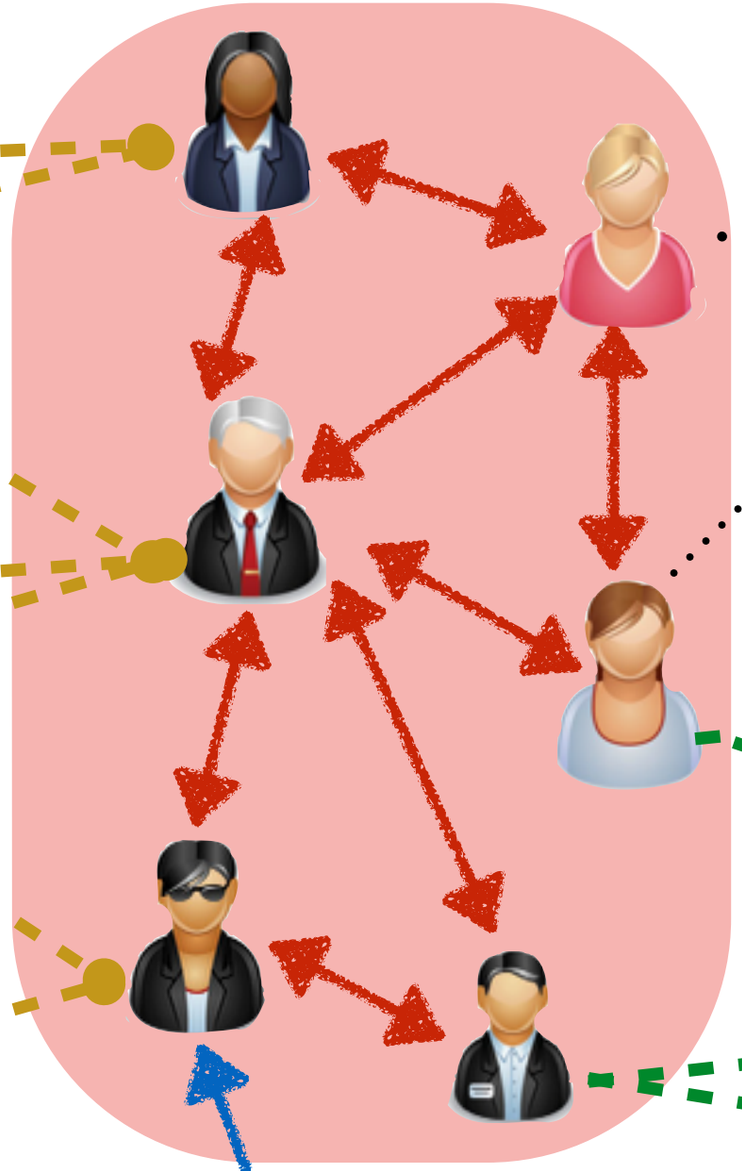
Solve Challenge 2:

Information Distribution Difference Problem

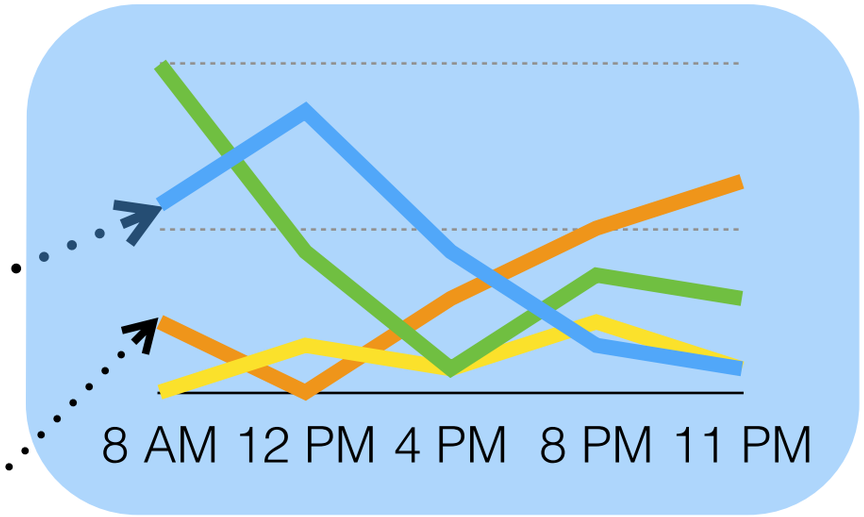
Locations



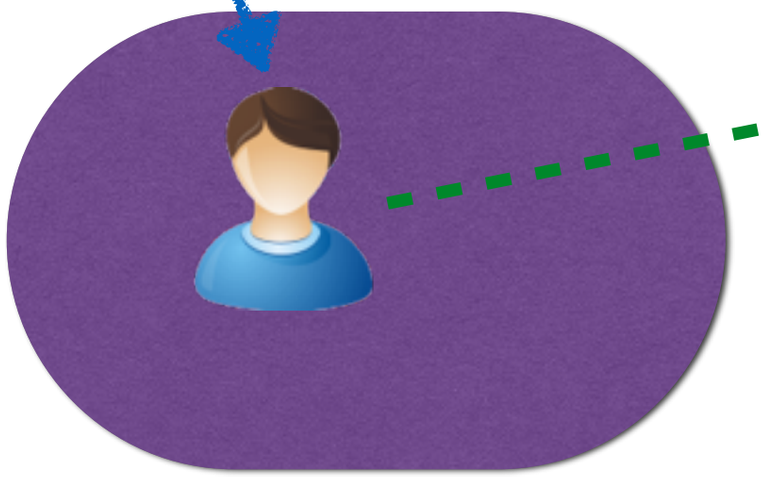
Social Links



Temporal Activities



Contents: Tweets



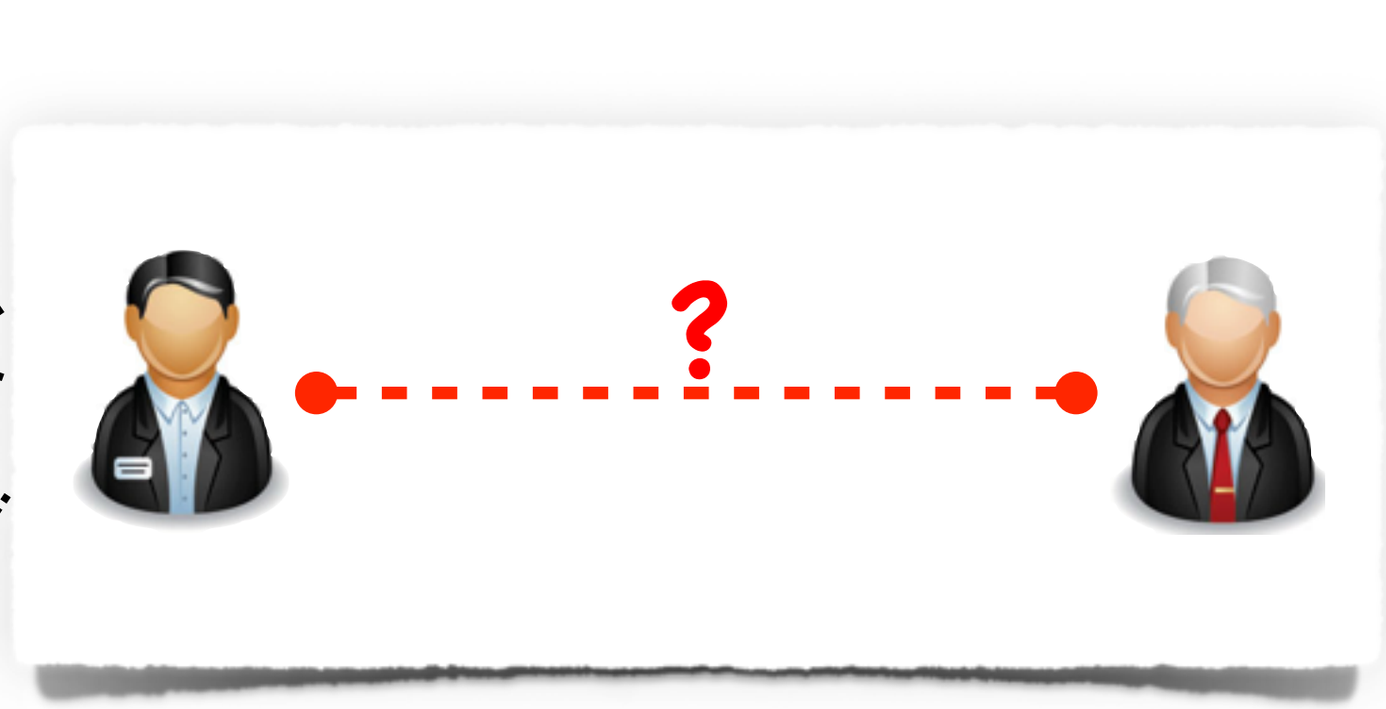
Extract Heterogeneous Features

Social

Spatial

Temporal

Content



training set

test set

8 AM 8 PM

Sheffield Neighbors

Near North Side

Near West Side

Chicago

Near South Side

South Loop

30

22.5

15

7.5

0

0

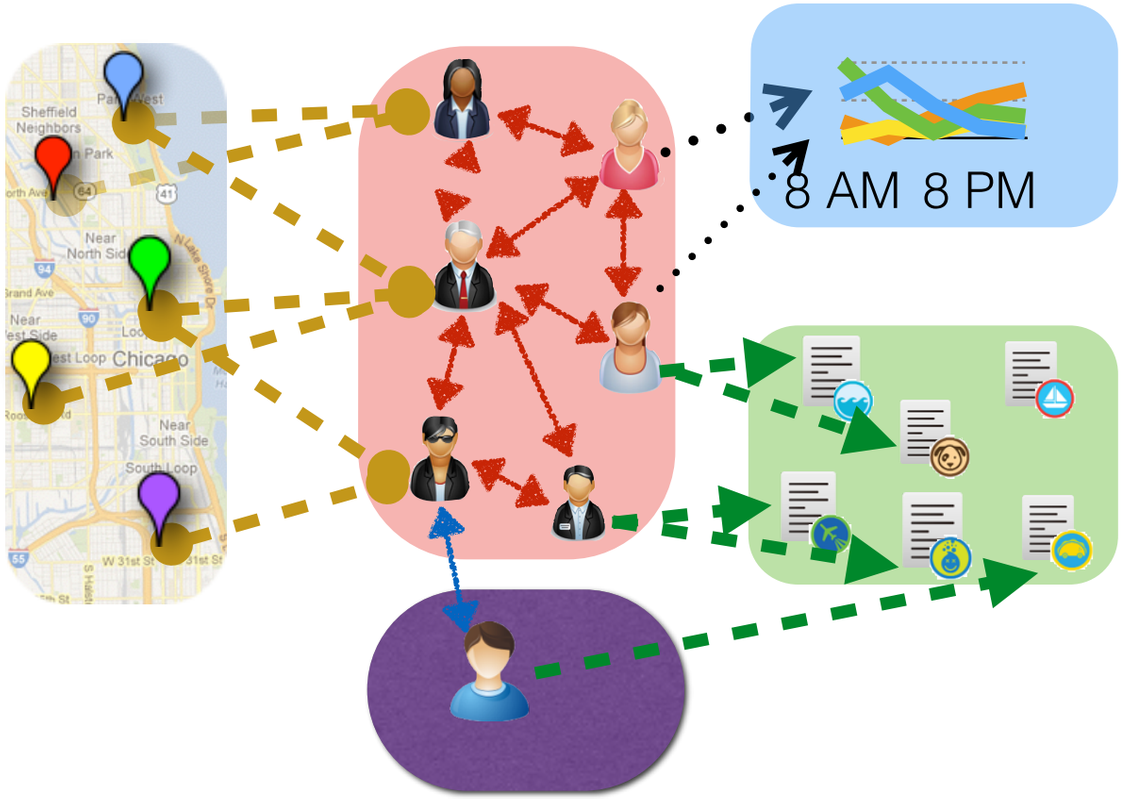
4

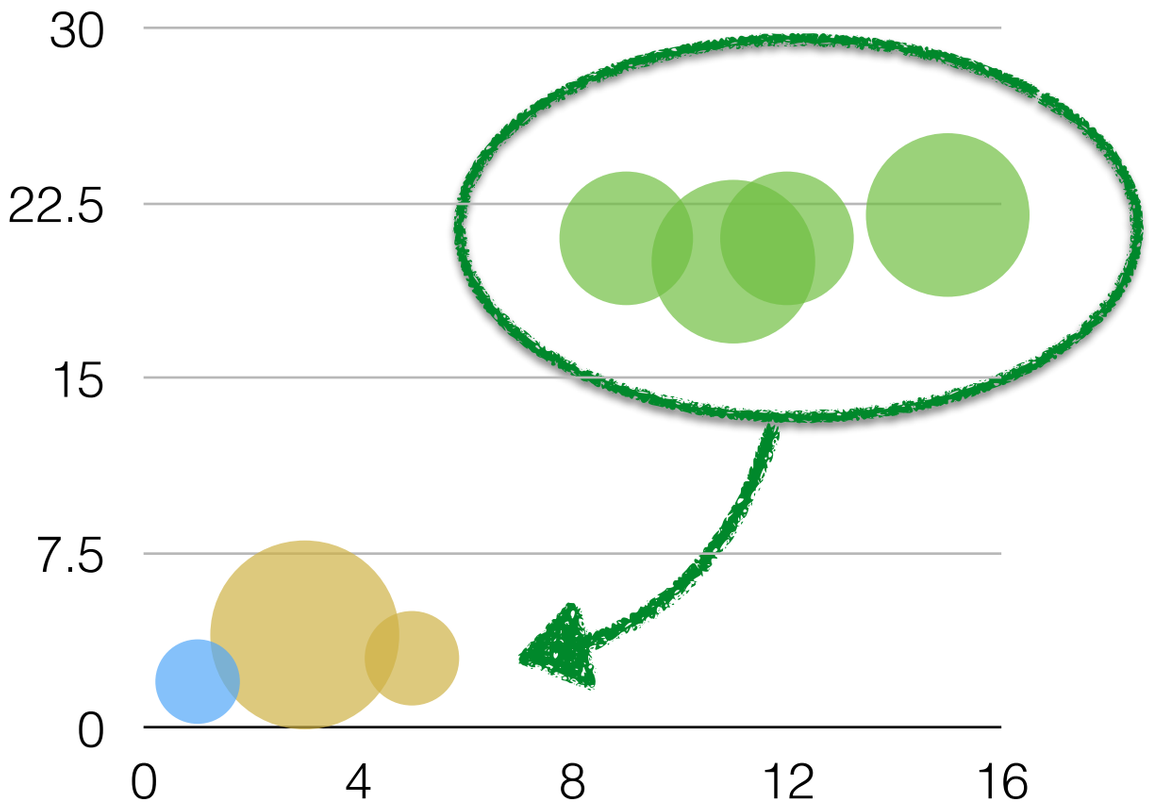
8

12

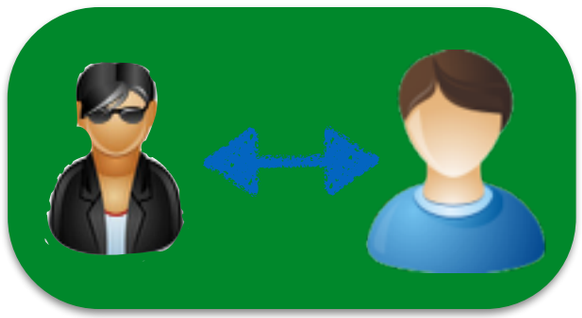
16

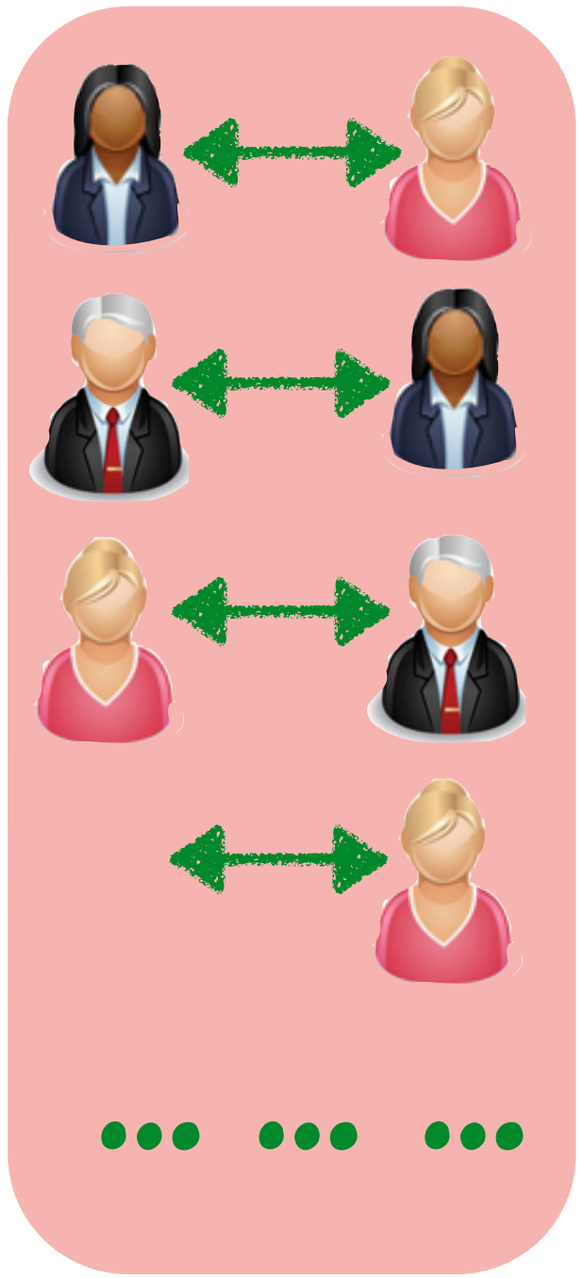
...



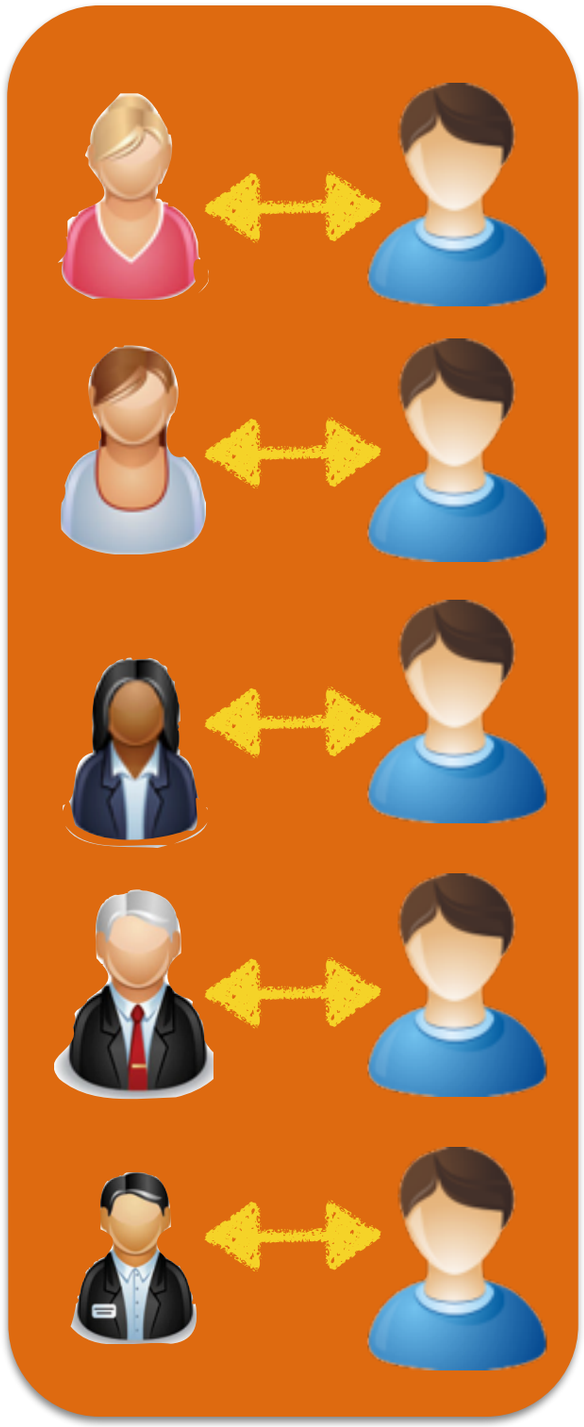


training set



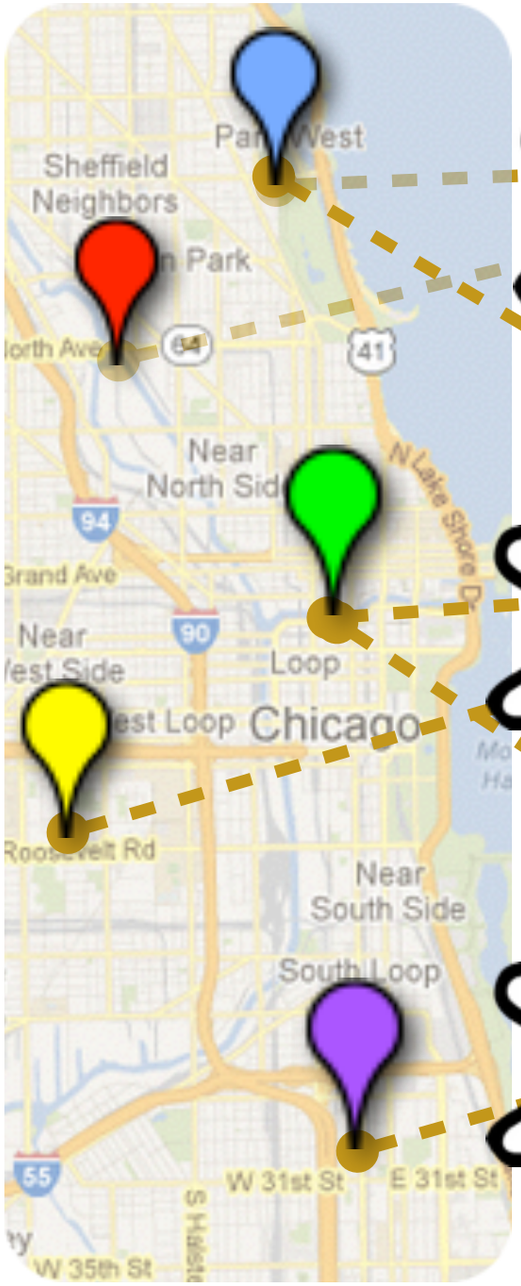


test set

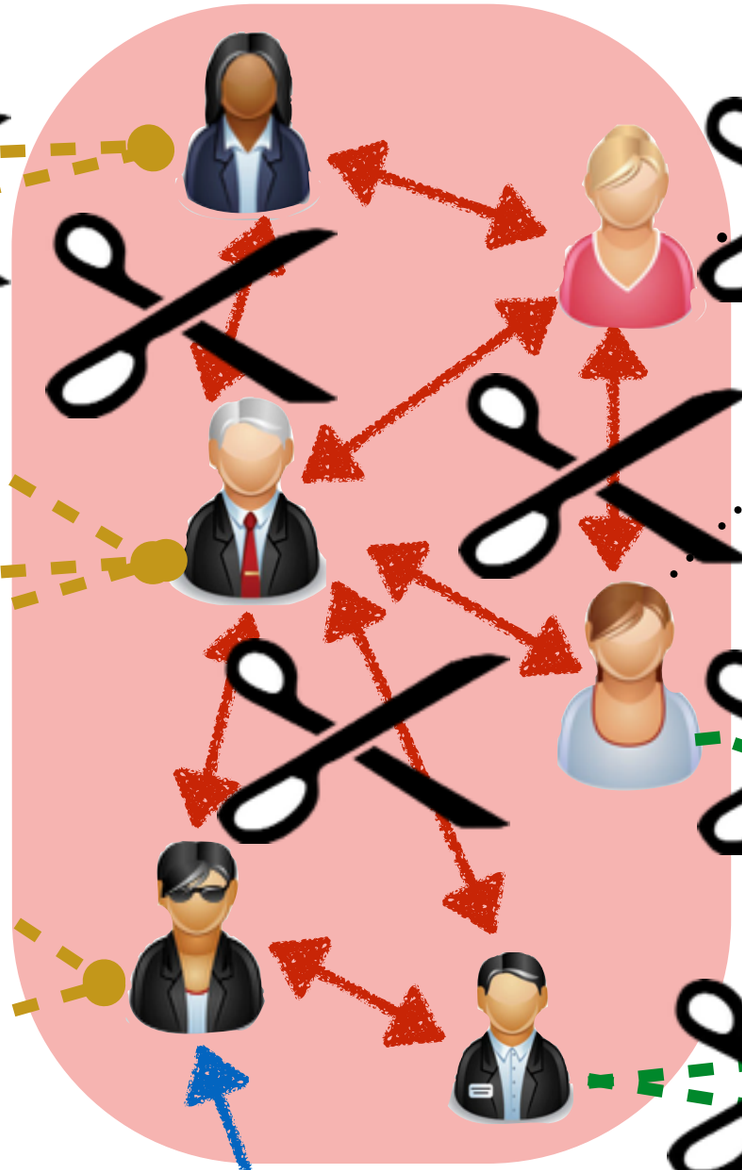


Personalized Random Sampling

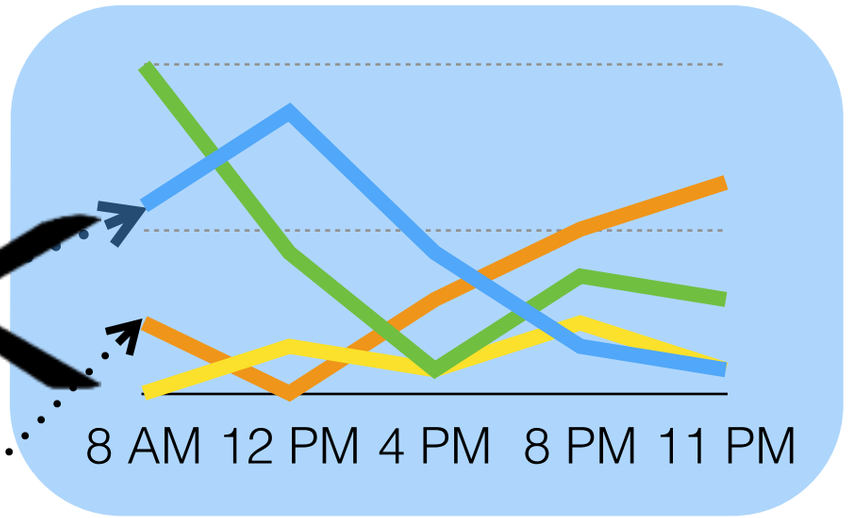
Locations



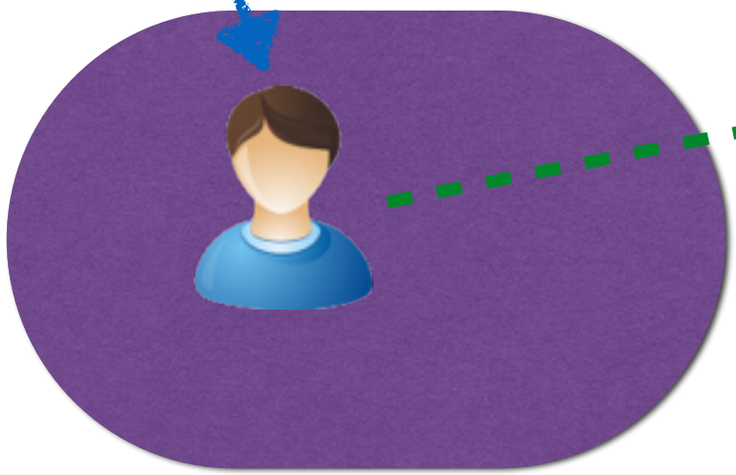
Social Links



Temporal Activities



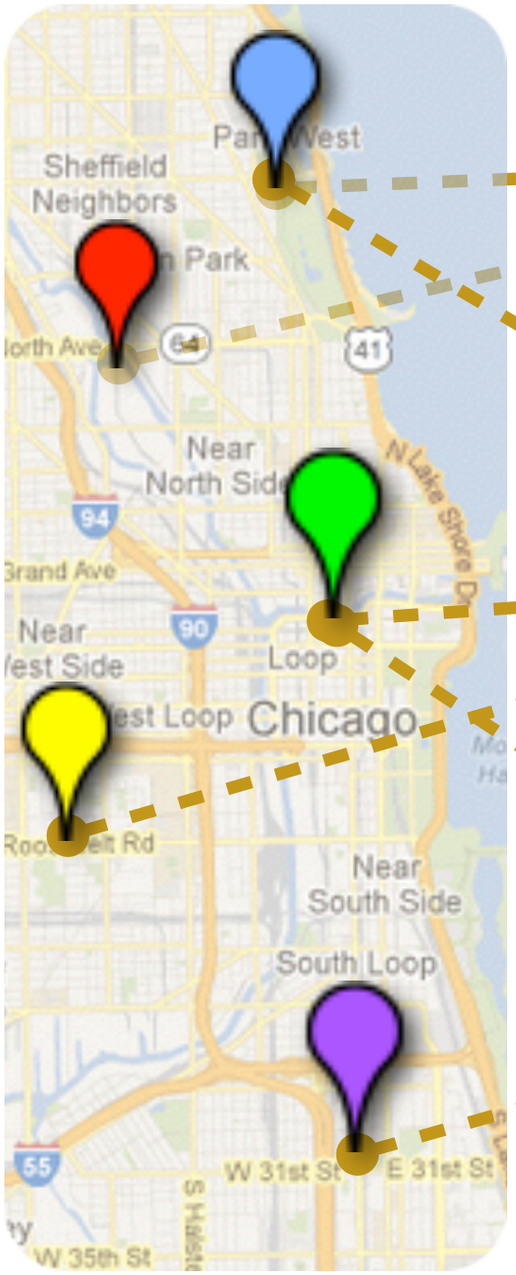
Contents: Tweets



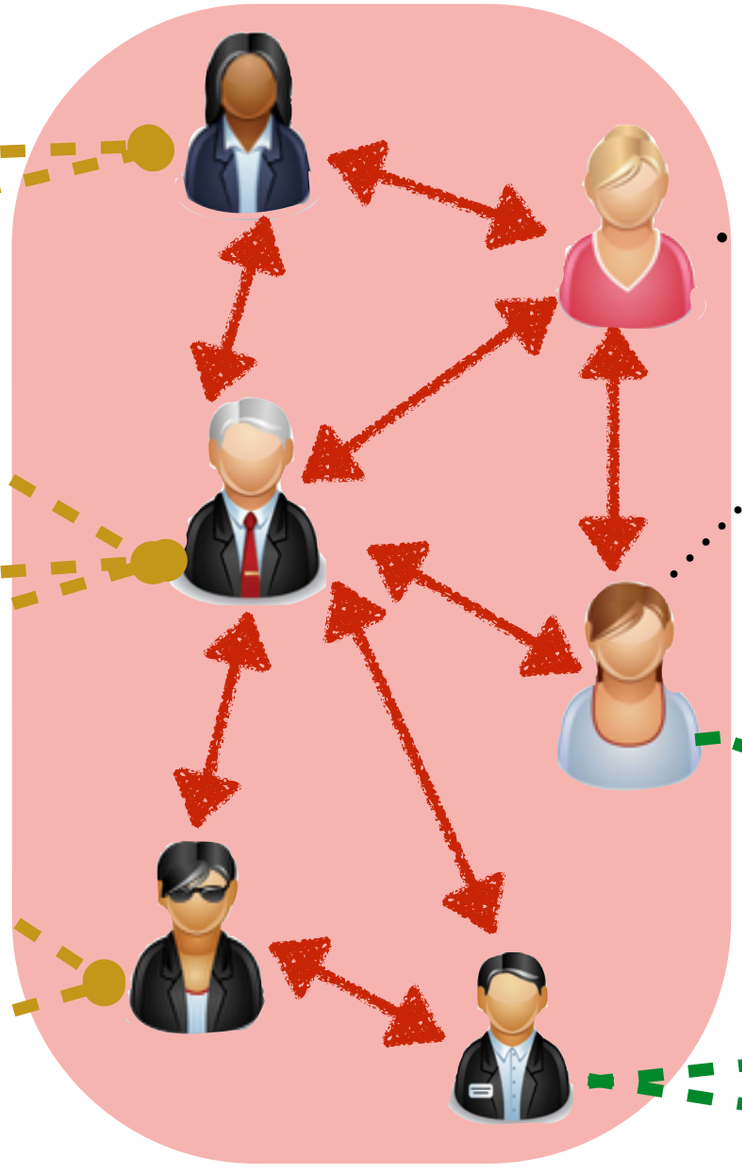
Solve Challenge 3:

Cold Start Problem

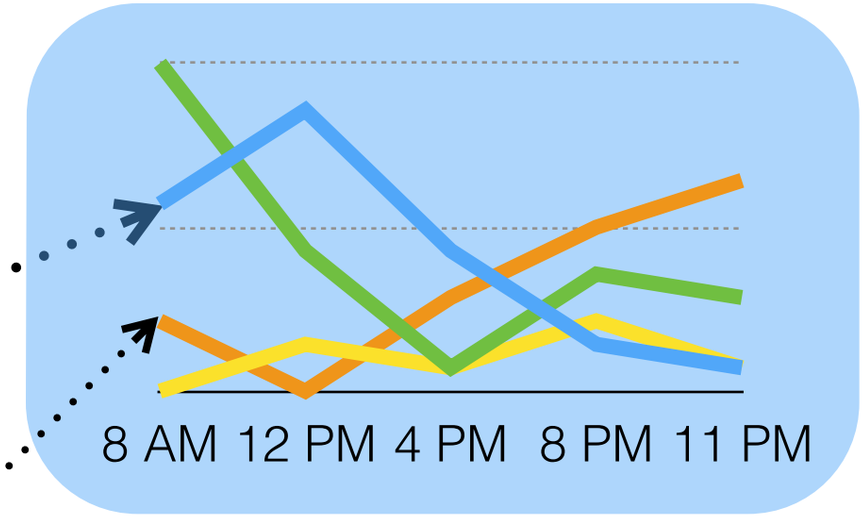
Locations



Social Links



Temporal Activities



Contents: Tweets


SMS
SMS

foursquare

I'm looking for...

Chicago, IL

Add Friends

Foursquare is better with your friends!

Find friends already using Foursquare via other networks around the web, or invite your friends using their email address

facebook

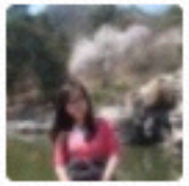
Gmail by Google

YAHOO!

Friends not on Foursquare? Invite them!

Invite your friends to Foursquare via Email.

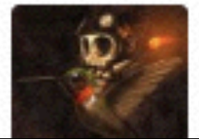
People you may know



ADD FRIEND

ADD FRIEND

ADD FRIEND



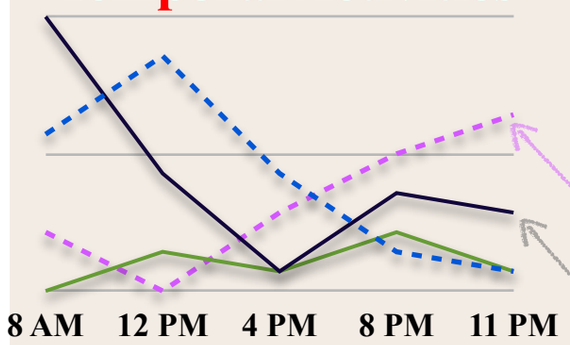
foursquare

twitter

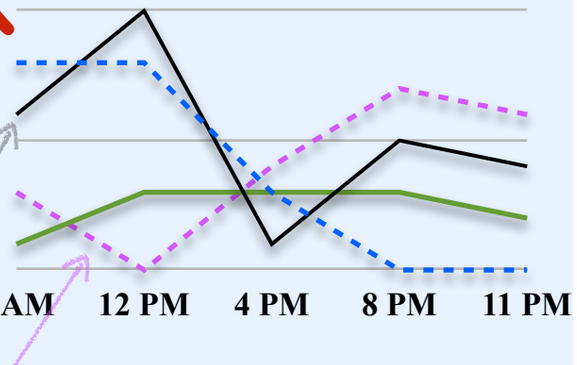
target network

source network

Temporal Activities



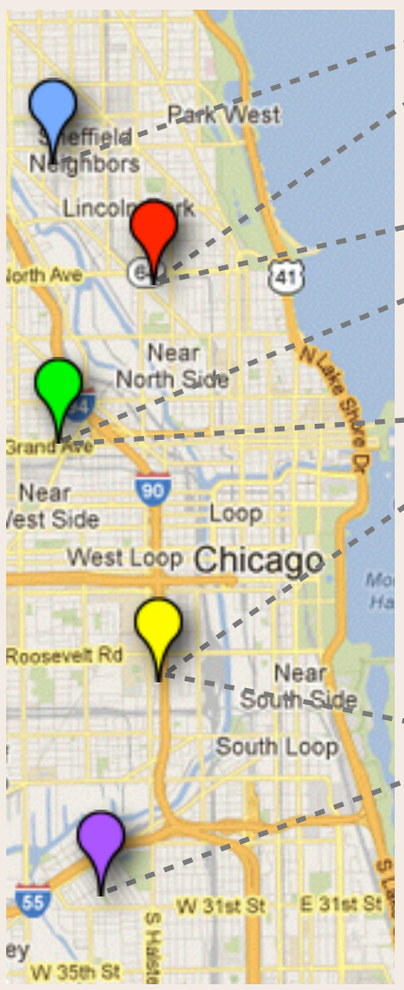
Temporal Activities



User Accounts

User Accounts

Locations



locate

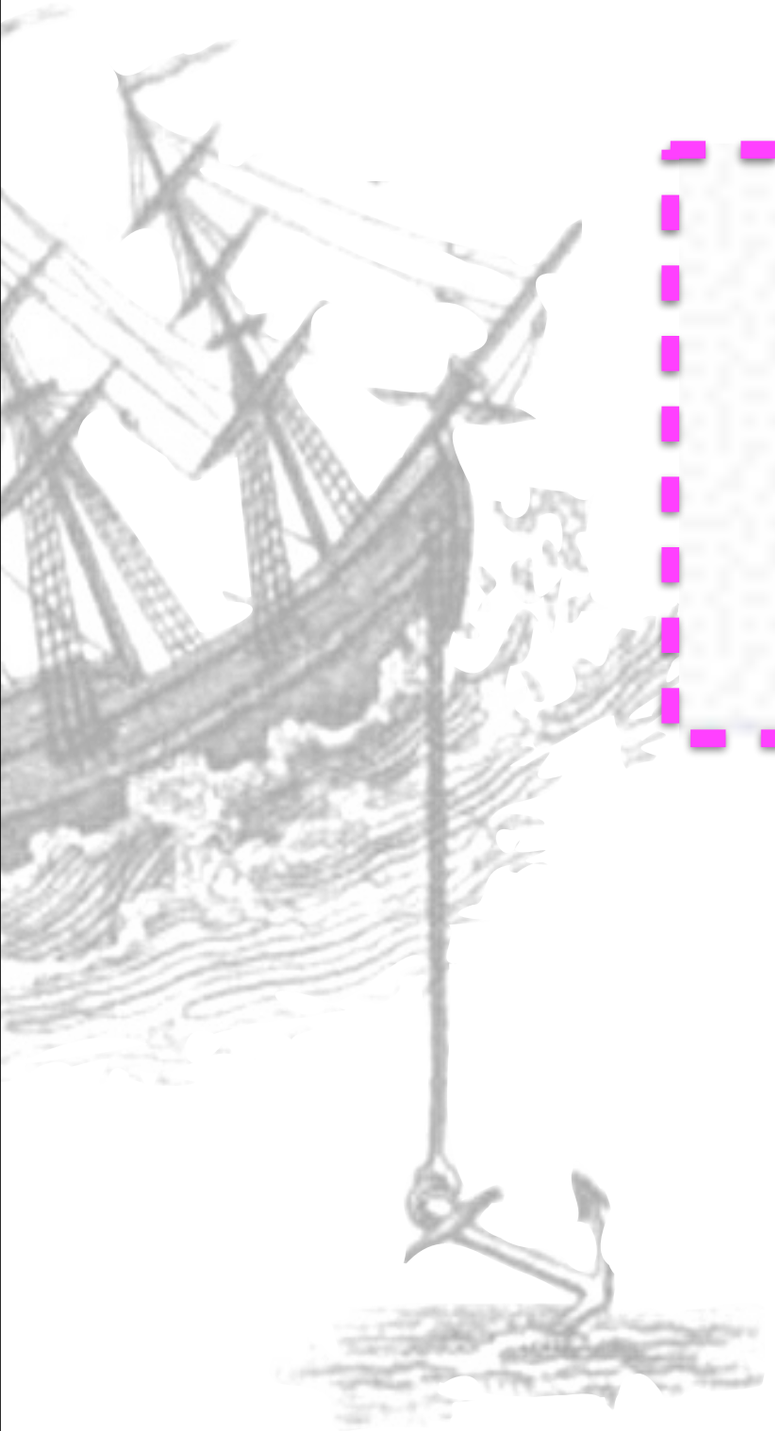
Locations

locate

Tips

Tweets

Anchor Links across Networks

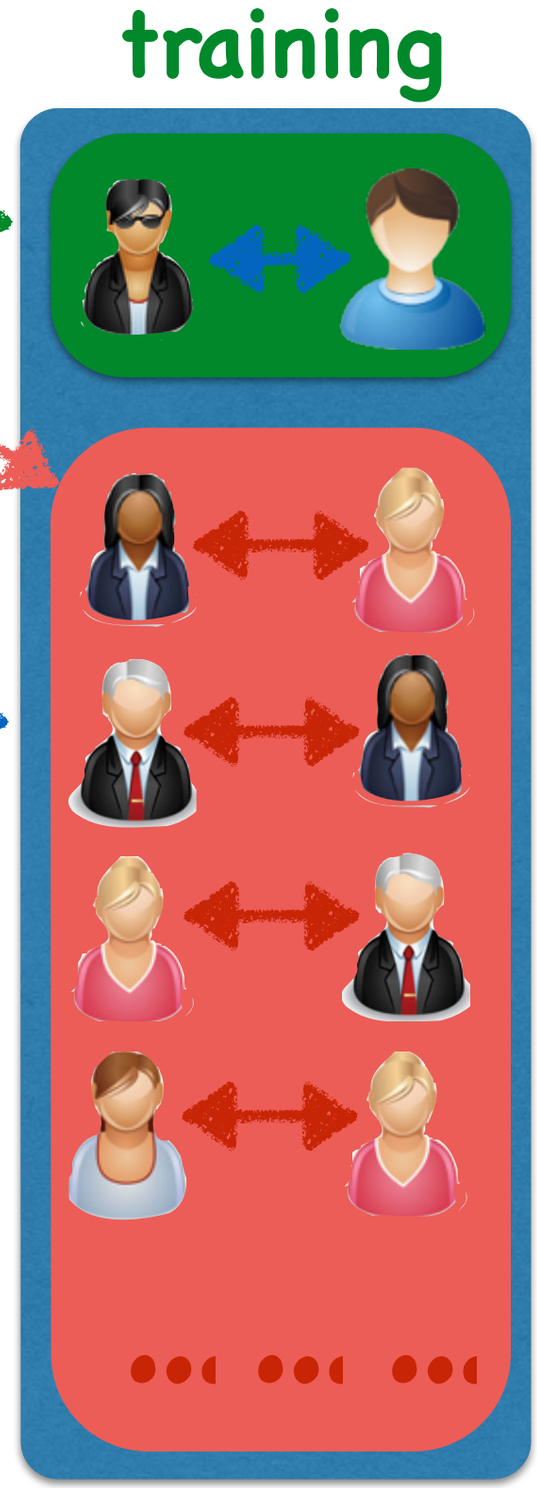


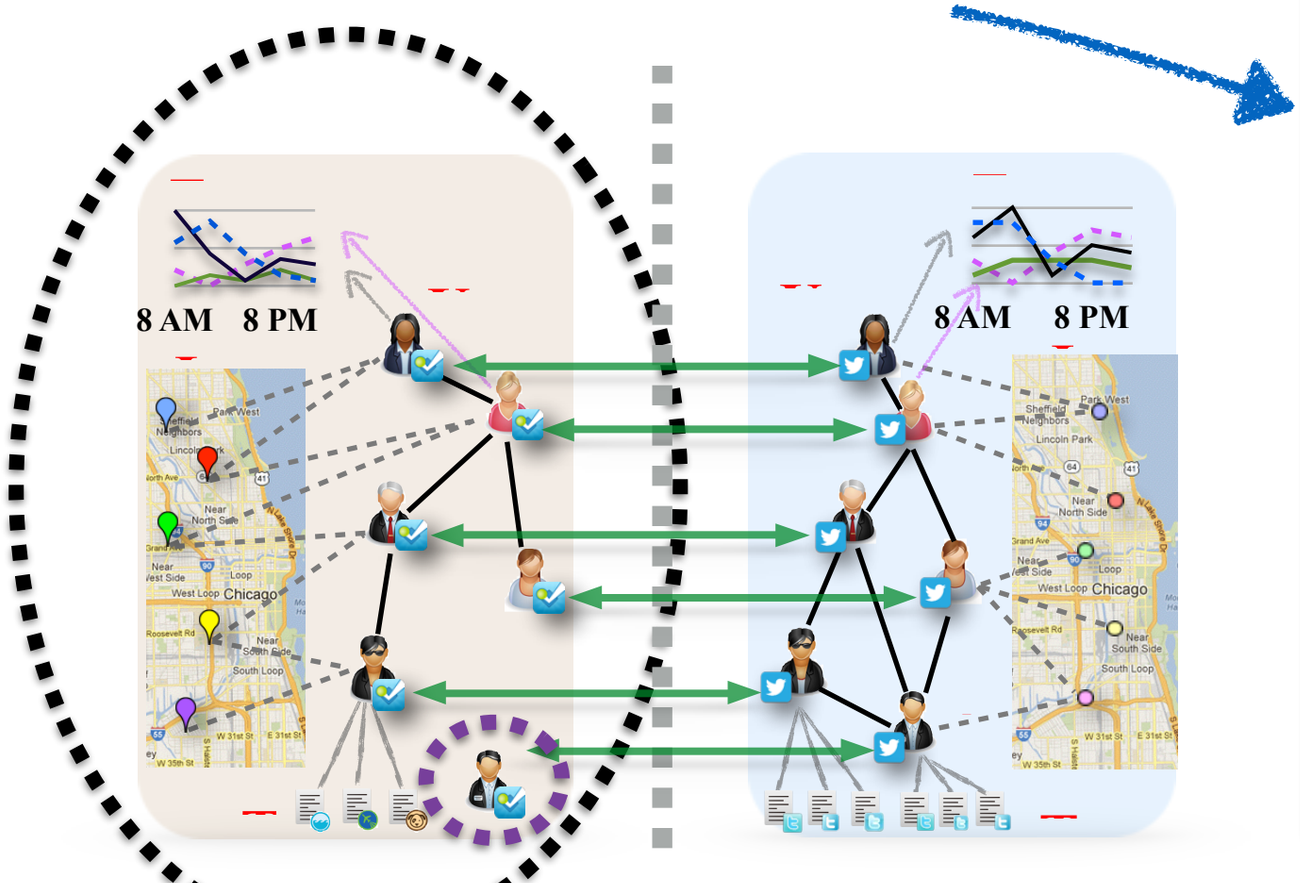
Data Sets

		network	
	property	**Twitter**	**Foursquare**
# node	user	5,223	5,392
	tweet/tip	9,490,707	48,756
	location	297,182	38,921
# link	friend/follow	164,920	31,312
	write	9,490,707	48,756
	locate	615,515	48,756

Comparison Methods

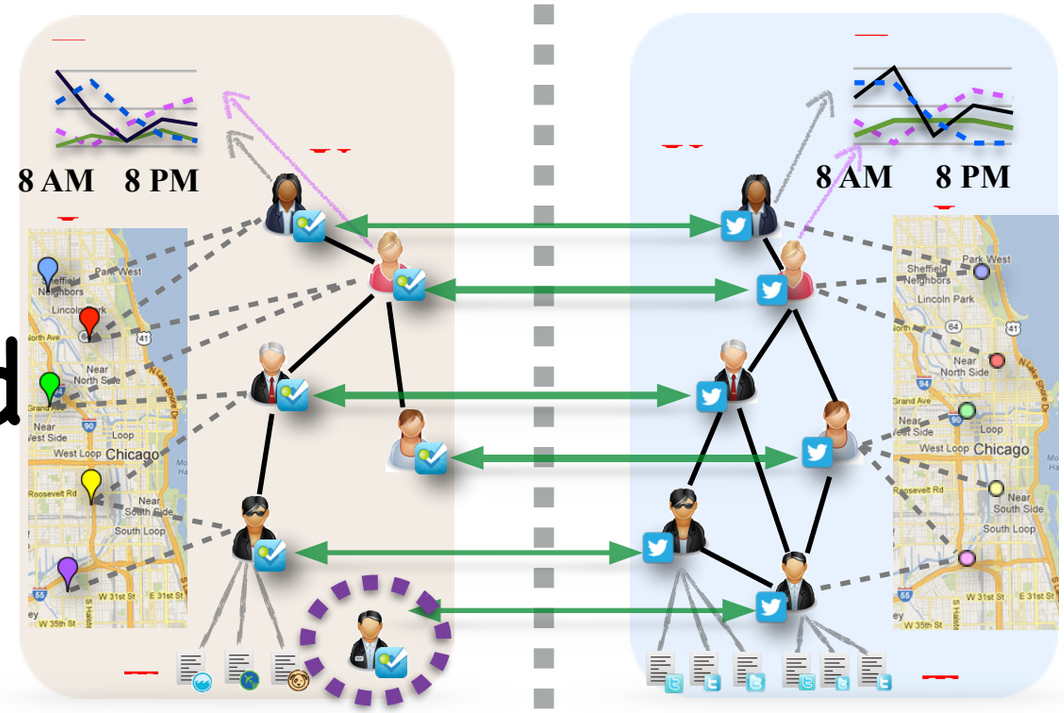
- Supervised Methods
  - New: New User's Social Links
  - Old: Old User's Social Links
  - TRAD: New + Old Users Social Links
  - Old-PS
  - TRAD-PS

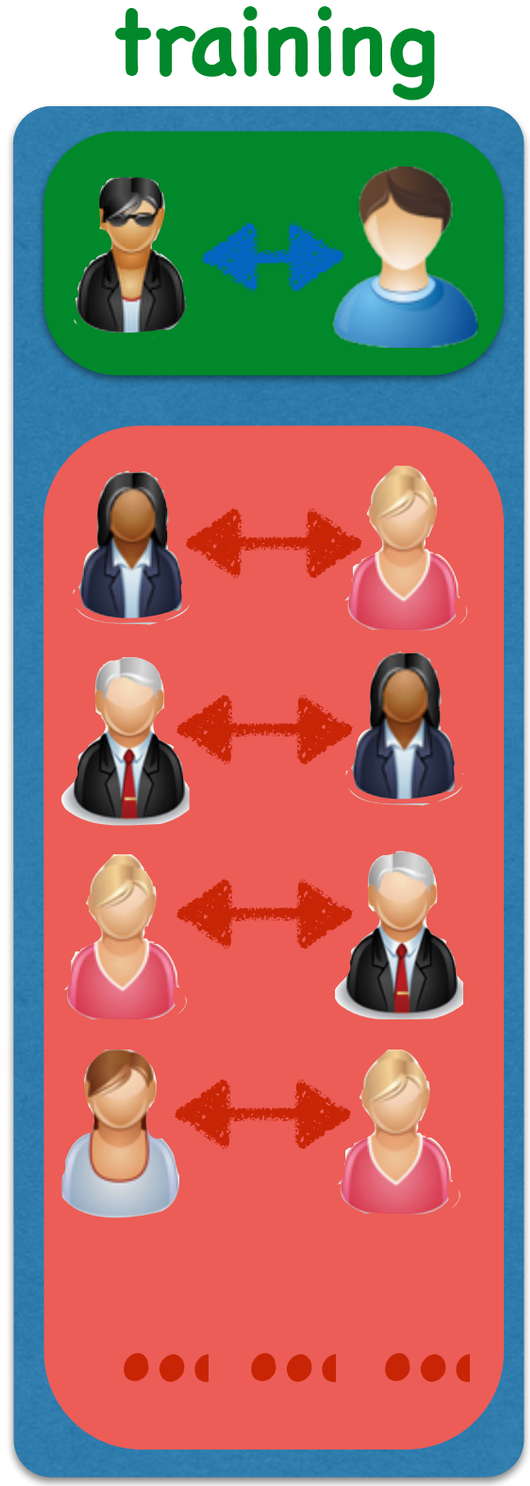




Comparison Methods

- Supervised Methods
  - Source
  - SCAN
  - SCAN-PS
- Unsupervised Methods
  - CN: Common Neighbor
  - JC: Jaccard Coefficient
  - AA: Adamic/Adar





Evaluation Methods

- AUC
- Accuracy

Experiment Results

target:

source:

degree of newness

measure	method	0.0	0.1	0.2
AUC	SCAN-PS	**0.783±0.009**	**0.839±0.008**	**0.864±0.013**
	SCAN	0.768±0.013	0.808±0.007	0.833±0.009
	SOURCE	0.761±0.008	0.768±0.015	0.800±0.014
	TRAD-PS	0.553±0.007	0.626±0.003	0.69±0.012
	OLD-PS	0.554±0.016	0.567±0.01	0.564±0.022
	TRAD	0.555±0.006	0.593±0.007	0.622±0.009
	OLD	0.550±0.008	0.510±0.010	0.527±0.008
	NEW	0.495±0.018	0.616±0.011	0.631±0.005
	CN	0.500±0.000	0.523±0.005	0.536±0.004
	JC	0.500±0.000	0.523±0.005	0.534±0.006
	AA	0.500±0.000	0.521±0.004	0.531±0.003
Acc.	SCAN-PS	**0.747±0.005**	**0.772±0.010**	**0.802±0.007**
	SCAN	0.732±0.014	0.746±0.008	0.763±0.010
	SOURCE	0.695±0.011	0.712±0.011	0.716±0.015
	TRAD-PS	0.506±0.004	0.600±0.006	0.610±0.009
	OLD-PS	0.506±0.002	0.504±0.002	0.505±0.004
	TRAD	0.506±0.002	0.524±0.006	0.540±0.004
	OLD	0.503±0.002	0.503±0.002	0.503±0.004
	NEW	0.478±0.010	0.563±0.009	0.581±0.004
	NAIVE	0.616±0.009	0.608±0.004	0.622±0.003

Experiment Results

target:

twitter

source:

foursquare

measure	method	0.0	0.1	0.2
AUC	SCAN-PS	**0.608±0.006**	**0.832±0.005**	**0.859±0.004**
	SCAN	0.602±0.005	0.788±0.005	0.827±0.003
	SOURCE	0.621±0.007	0.736±0.005	0.734±0.005
	TRAD-PS	0.526±0.004	0.772±0.006	0.785±0.002
	OLD-PS	0.530±0.003	0.680±0.007	0.653±0.006
	TRAD	0.456±0.003	0.697±0.007	0.772±0.004
	OLD	0.423±0.002	0.519±0.004	0.528±0.005
	NEW	0.492±0.013	0.766±0.008	0.788±0.003
	CN	0.500±0.000	0.731±0.006	0.786±0.001
	JC	0.500±0.000	0.716±0.007	0.760±0.002
	AA	0.500±0.000	0.728±0.005	0.782±0.002
Acc.	SCAN-PS	**0.588±0.001**	**0.769±0.004**	**0.793±0.005**
	SCAN	0.582±0.004	0.685±0.007	0.715±0.004
	SOURCE	0.573±0.006	0.669±0.005	0.676±0.003
	TRAD-PS	0.505±0.002	0.710±0.001	0.705±0.005
	OLD-PS	0.515±0.003	0.501±0.013	0.503±0.002
	TRAD	0.503±0.002	0.545±0.005	0.625±0.002
	OLD	0.516±0.006	0.500±0.002	0.513±0.001
	NEW	0.488±0.008	0.661±0.006	0.707±0.003
	NAIVE	0.552±0.003	0.552±0.002	0.553±0.002

Summary

- Problem Studied
  - **Social link prediction for new users**
- Novelty
  - within-network information transfer
  - information distribution difference
  - cross aligned network information transfer
  - cold start problem

Q&A