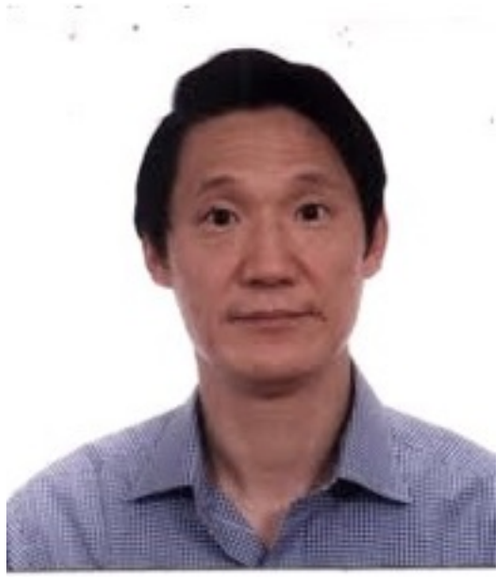
# Inferring Anchor Links across Multiple Heterogeneous Social Networks



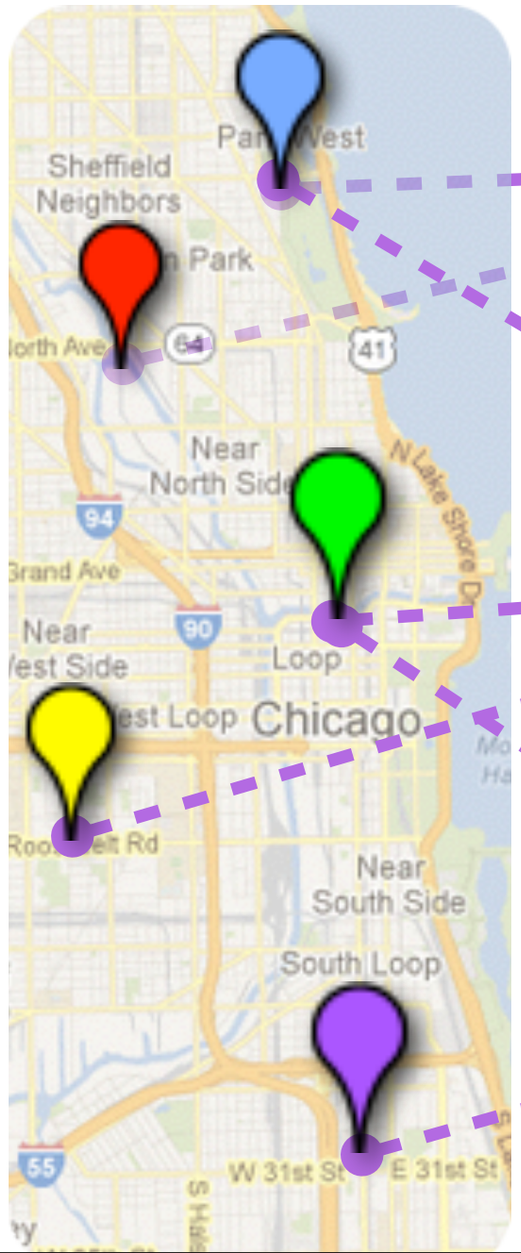
**Xiangnan Kong** Jiawei Zhang Philip S. Yu

University of Illinois at Chicago

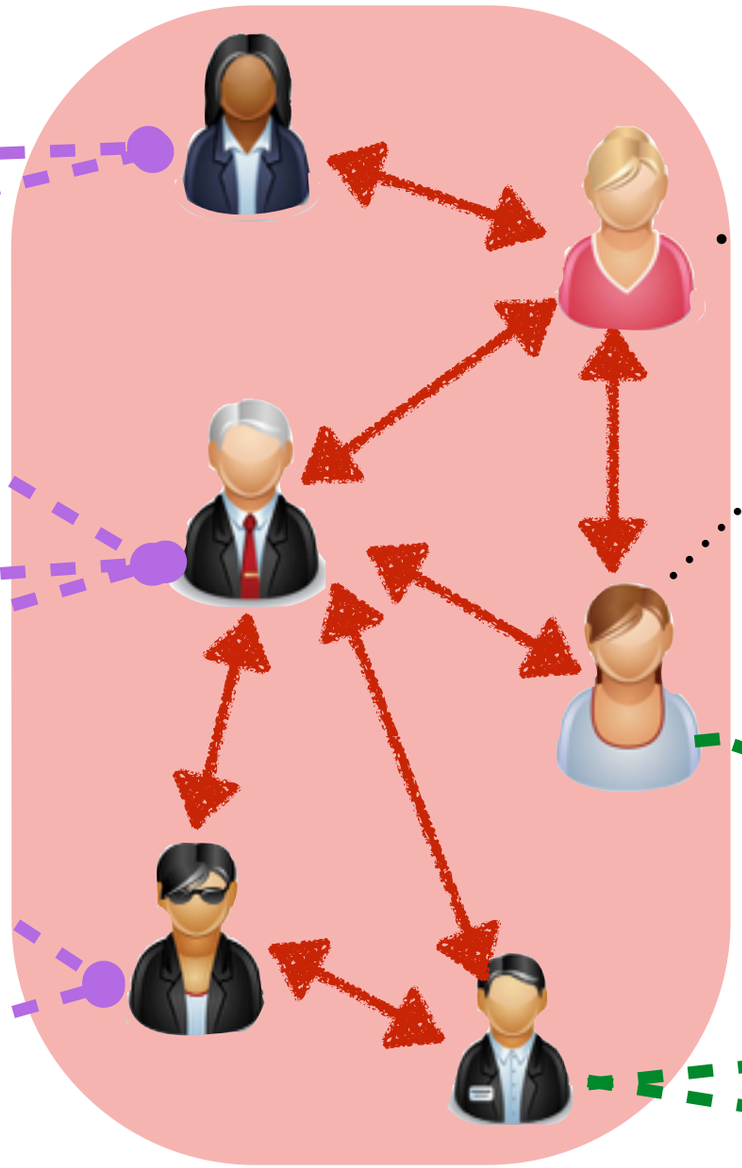
# Social Network:

## Who Where What When

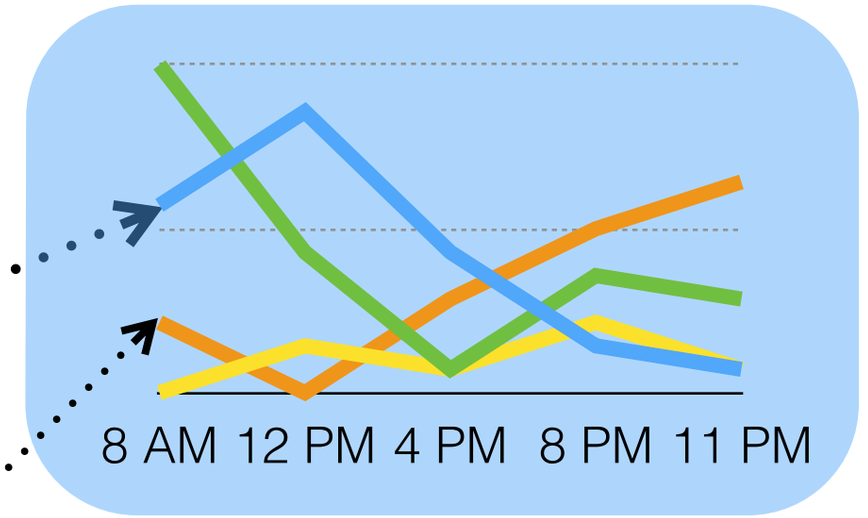
### Locations




### Social Links



### Temporal Activities




### Contents: Tweets

# Big Data

foursquare

twitter

Temporal Activities

8 AM 12 PM 4 PM 8 PM 11 PM

User Accounts

Locations

Tips

Temporal Activities

8 AM 12 PM 4 PM 8 PM 11 PM

User Accounts

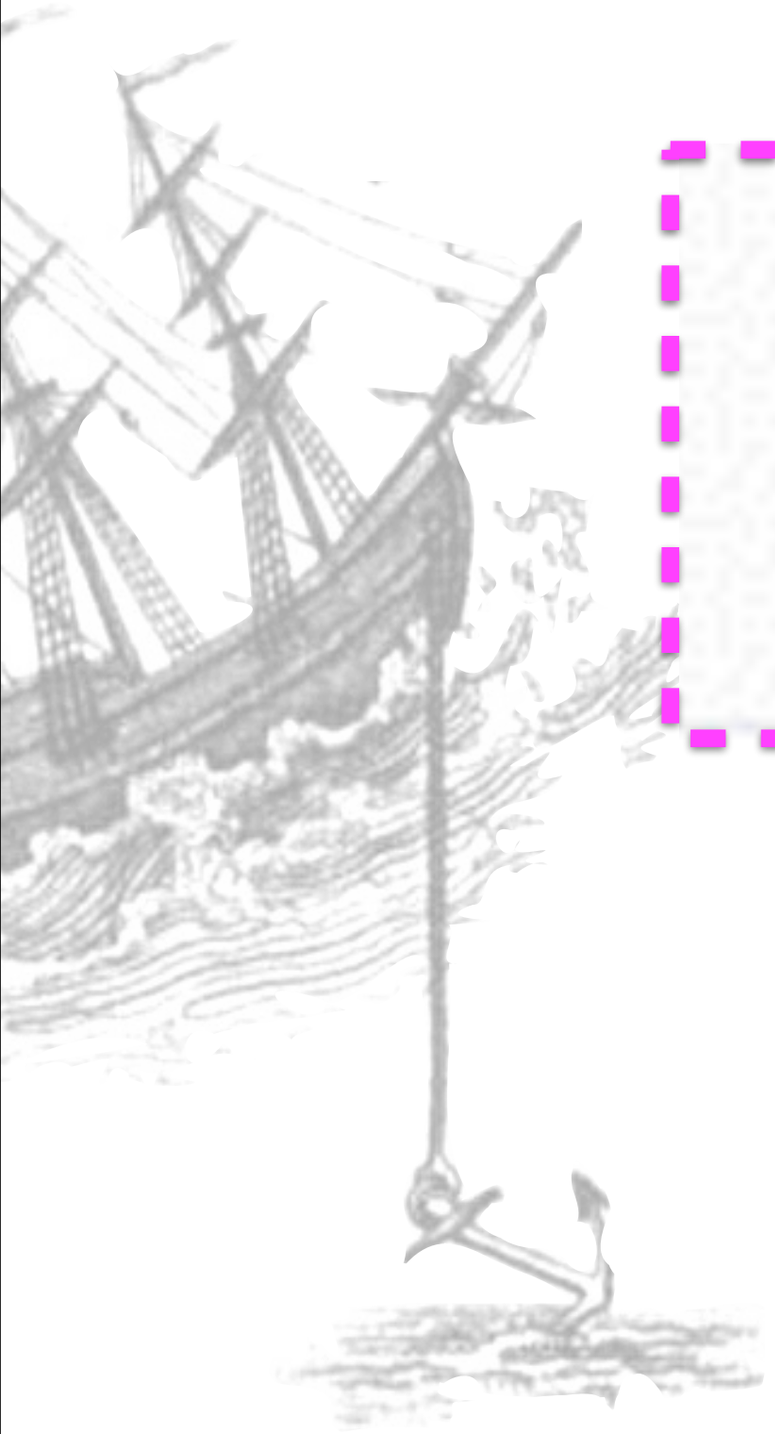
Locations

locate

locate

Tweets

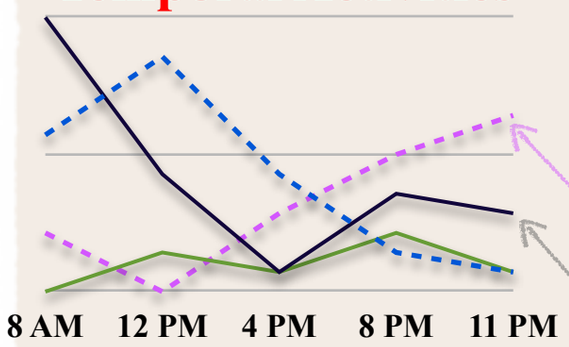
# Anchor Links across Networks



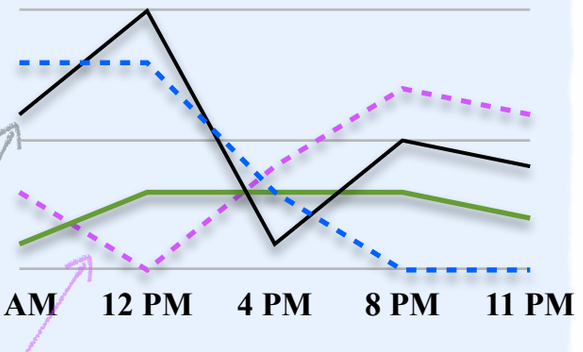
foursquare

twitter

Temporal Activities




Temporal Activities

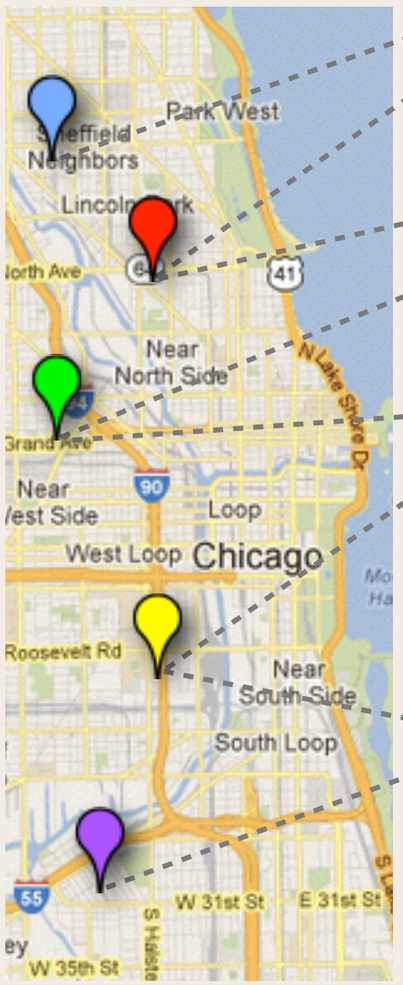



User Accounts

User Accounts

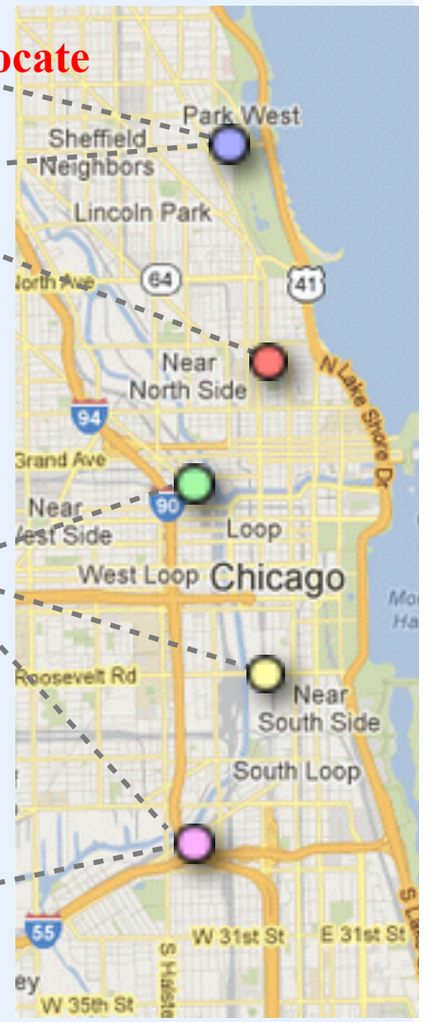
anchor links

Locations




locate

Locations




locate

?

?

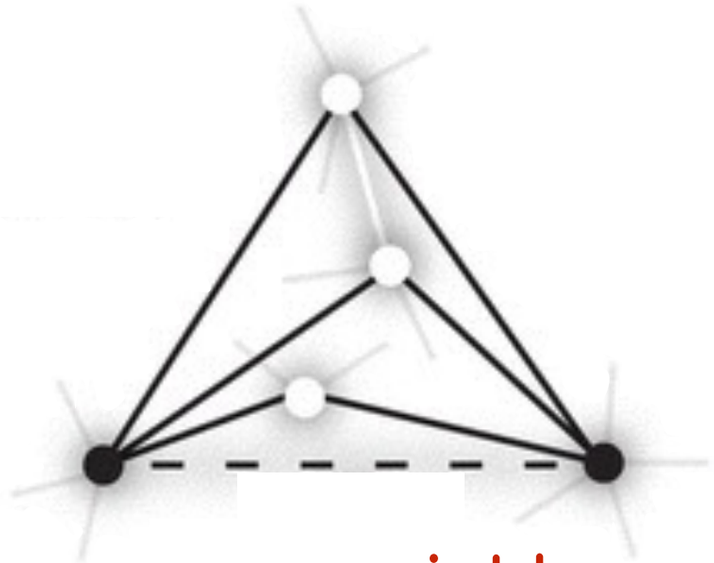
Tips

Tweets

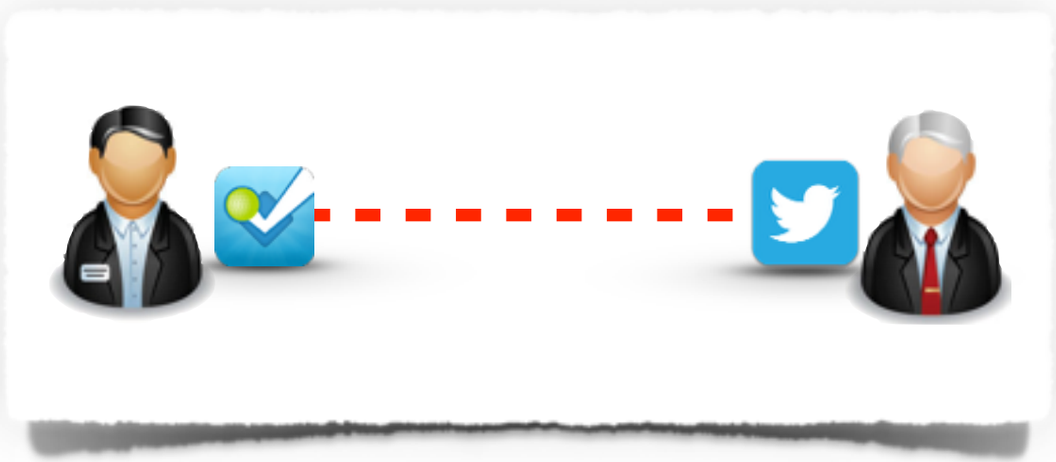
# Challenges of Supervised Anchor Link Prediction

## Training / Learning:

### Lack of Features

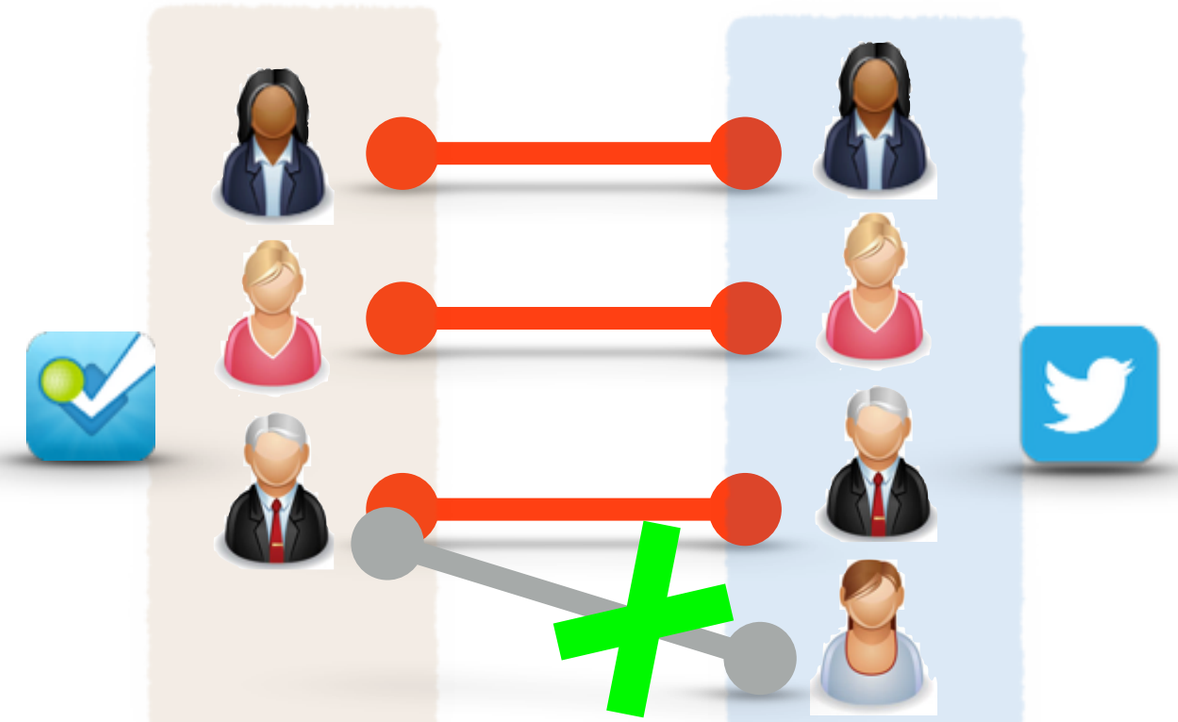


# common neighbors



## Testing /Inference:

### one-one Constraint



# Solve Challenge 1:

# Training/Learning

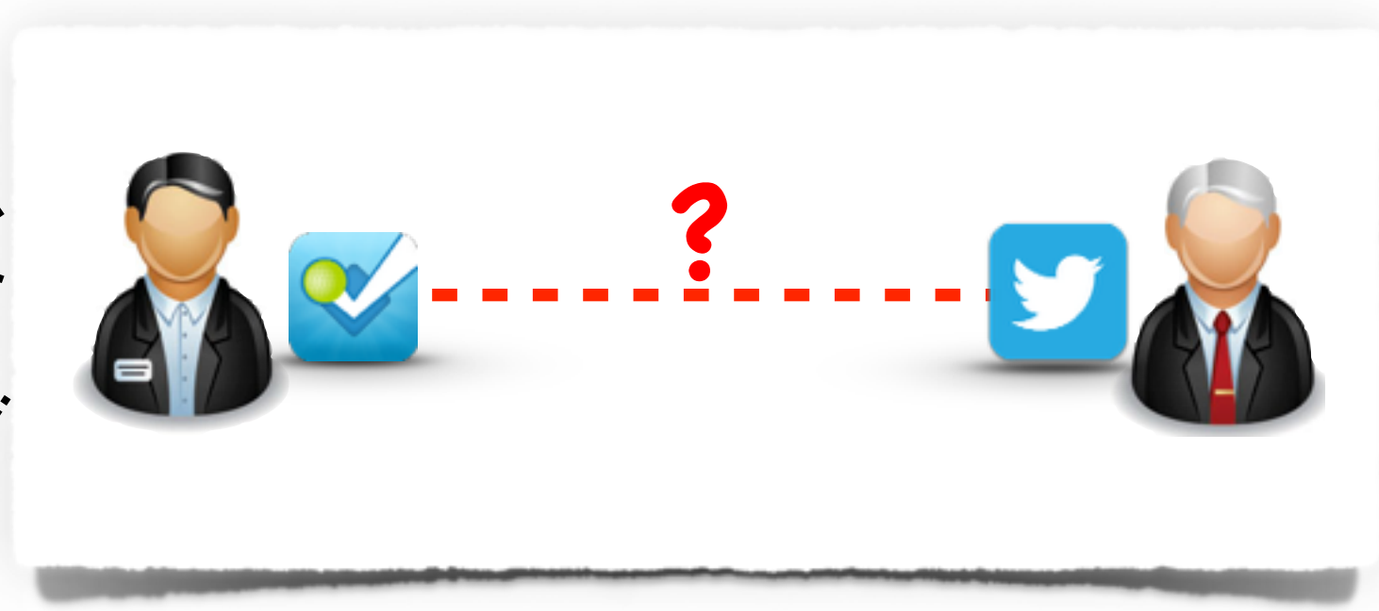
# Extract Heterogeneous Cross-Network Features

Social

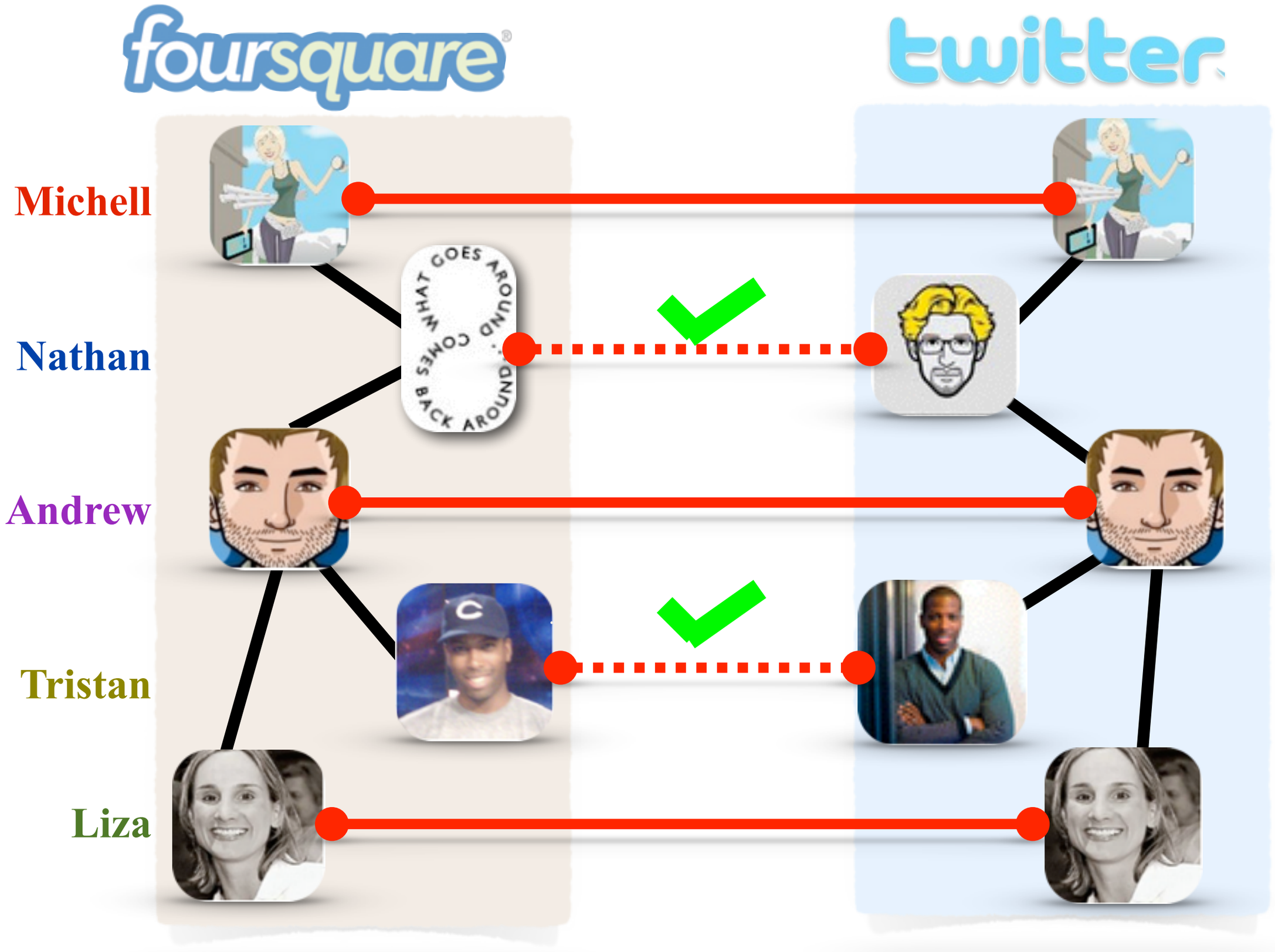
Spatial

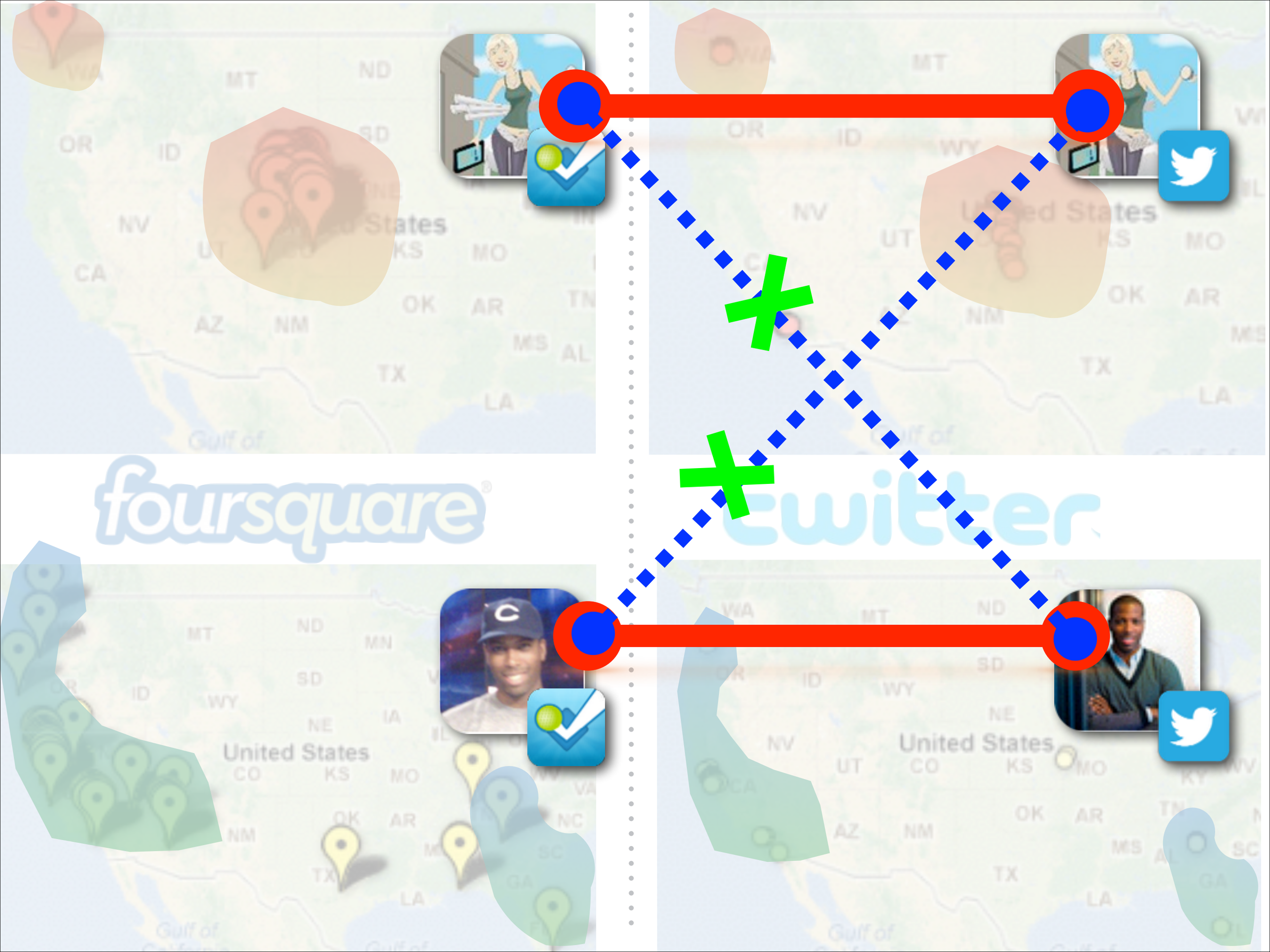
Temporal

Content



# Social



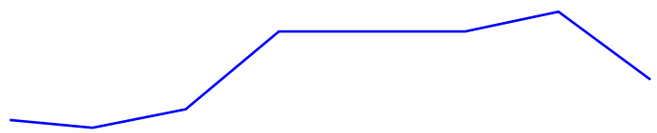

foursquare
twitter
United States
United States

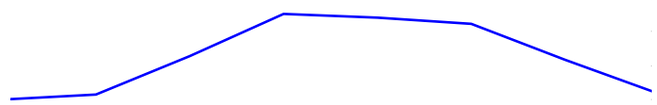
# Temporal

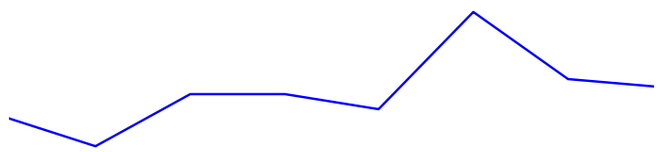
# User Activities

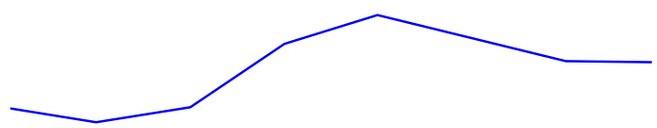
foursquare

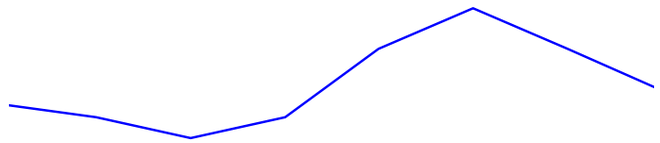
twitter

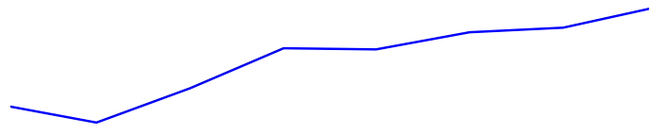


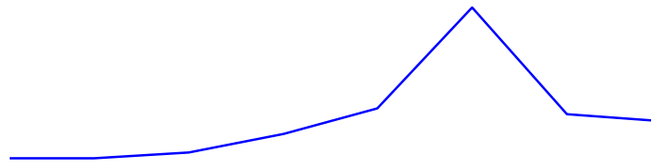


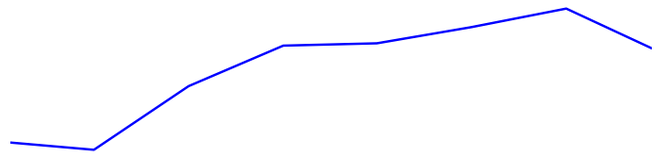


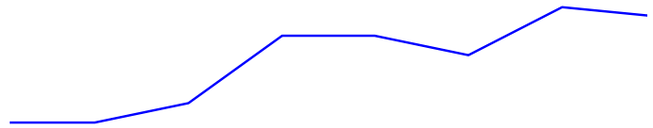


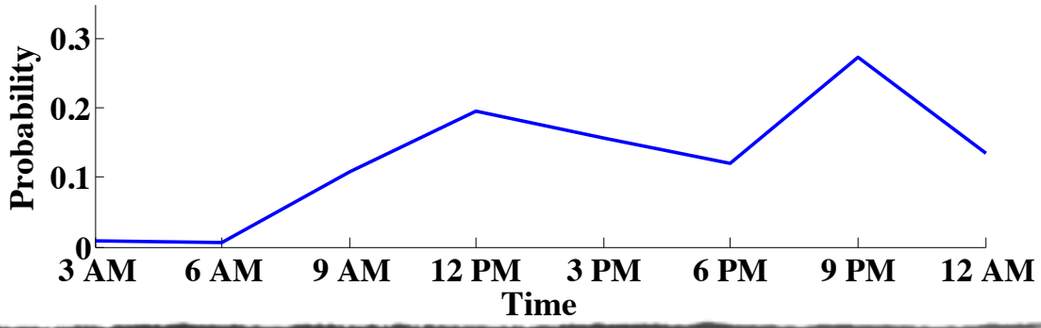












# Content

# Choice of Words

art (65,2), style (16,3)
audit (3,2), grill (19,2)

happy (27,5), enjoy (9,4)
week (18,4), shows (6,6)

awsm (2,3), kids (20,3)
red (61,3), open (11,4)

ask (6,5), coffee (8,3)
mochi (1,3), hangout (5,2)

win (19,4), amazing (55,5)
awesome (51,4), please (9,4)

awsm

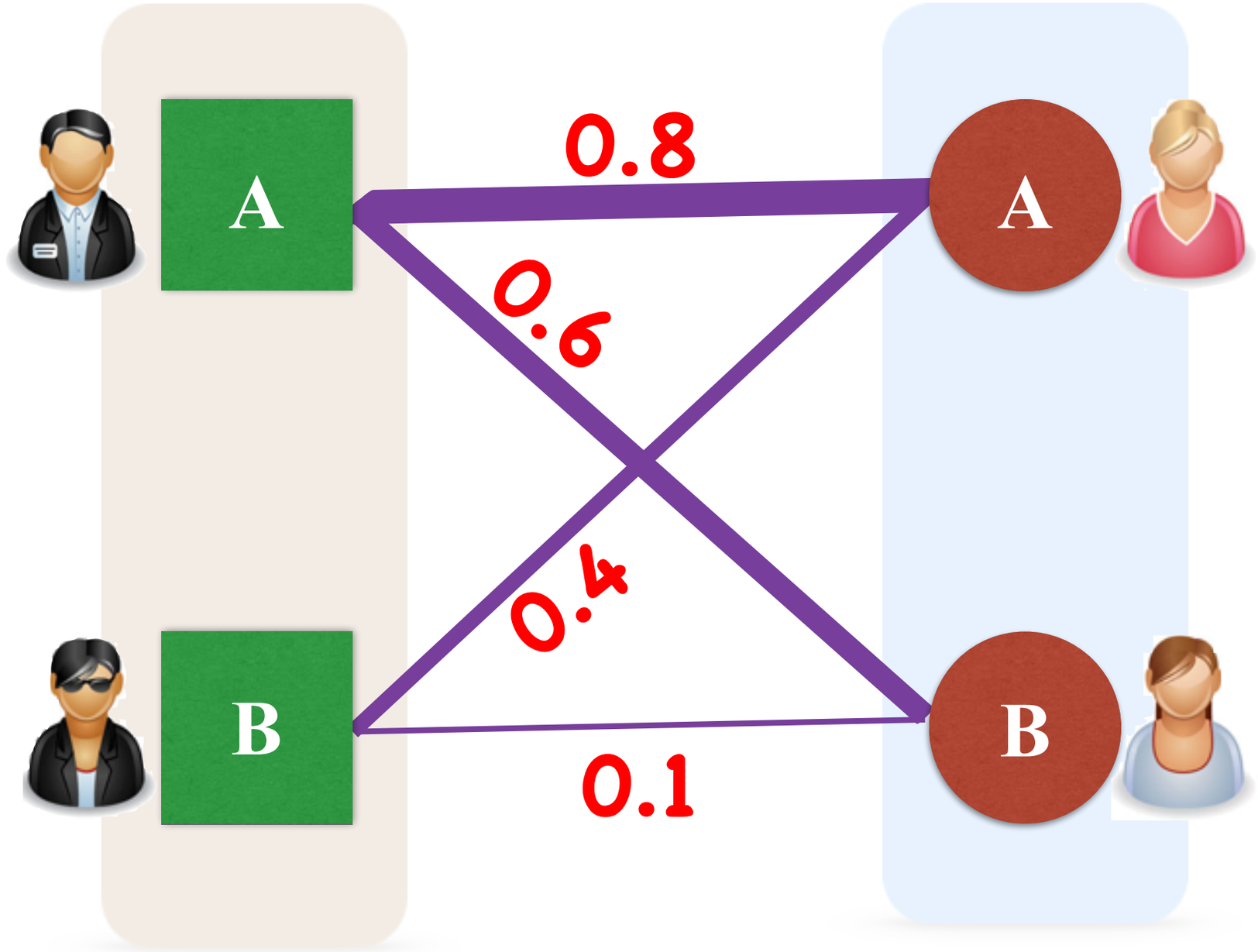
=

awesome

# Solve Challenge 11:

# Inference w.r.t. Constraints

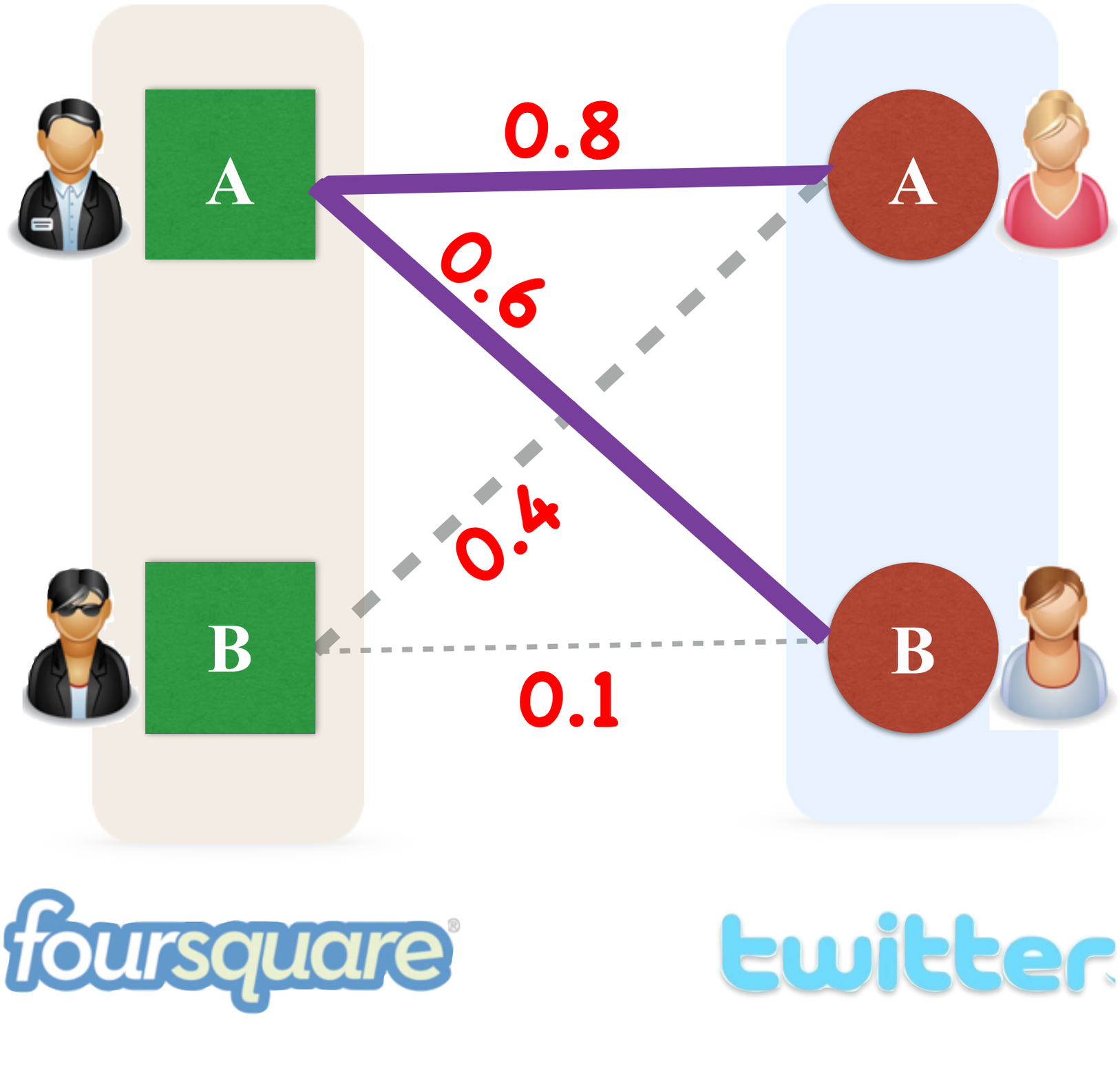
# Predicted Scores



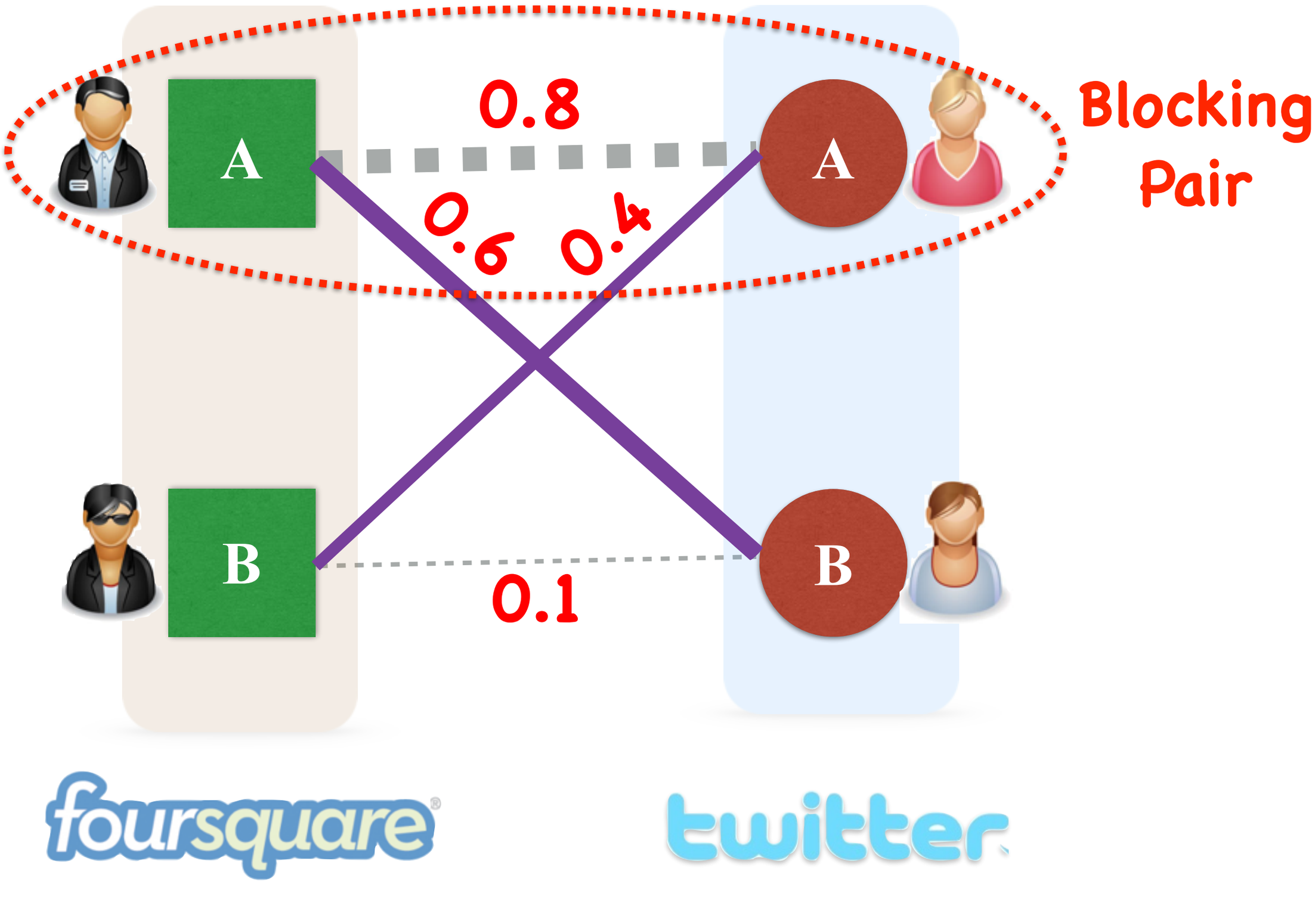

foursquare

twitter

# Link Prediction

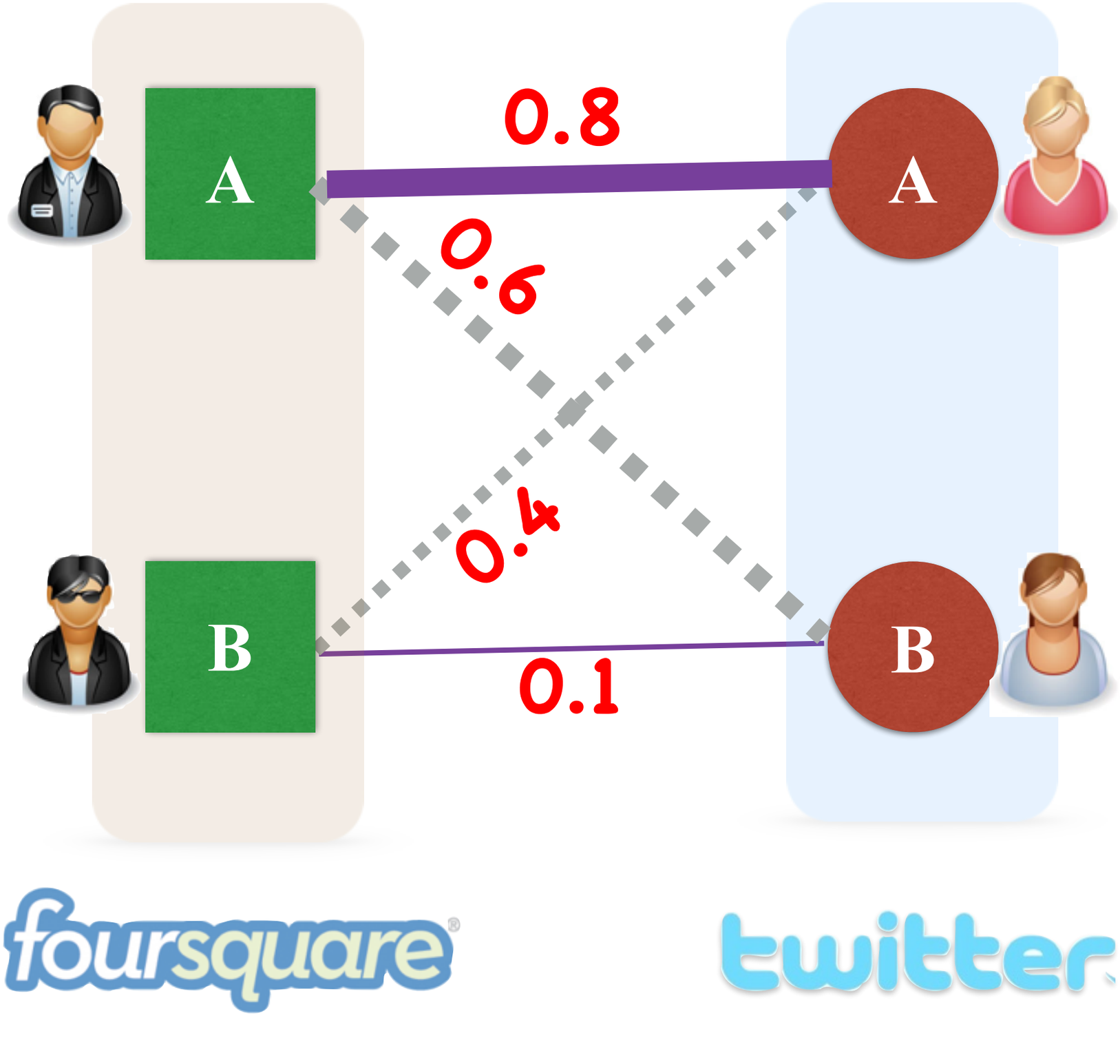


# Max Sum of Scores w.r.t. Constraints

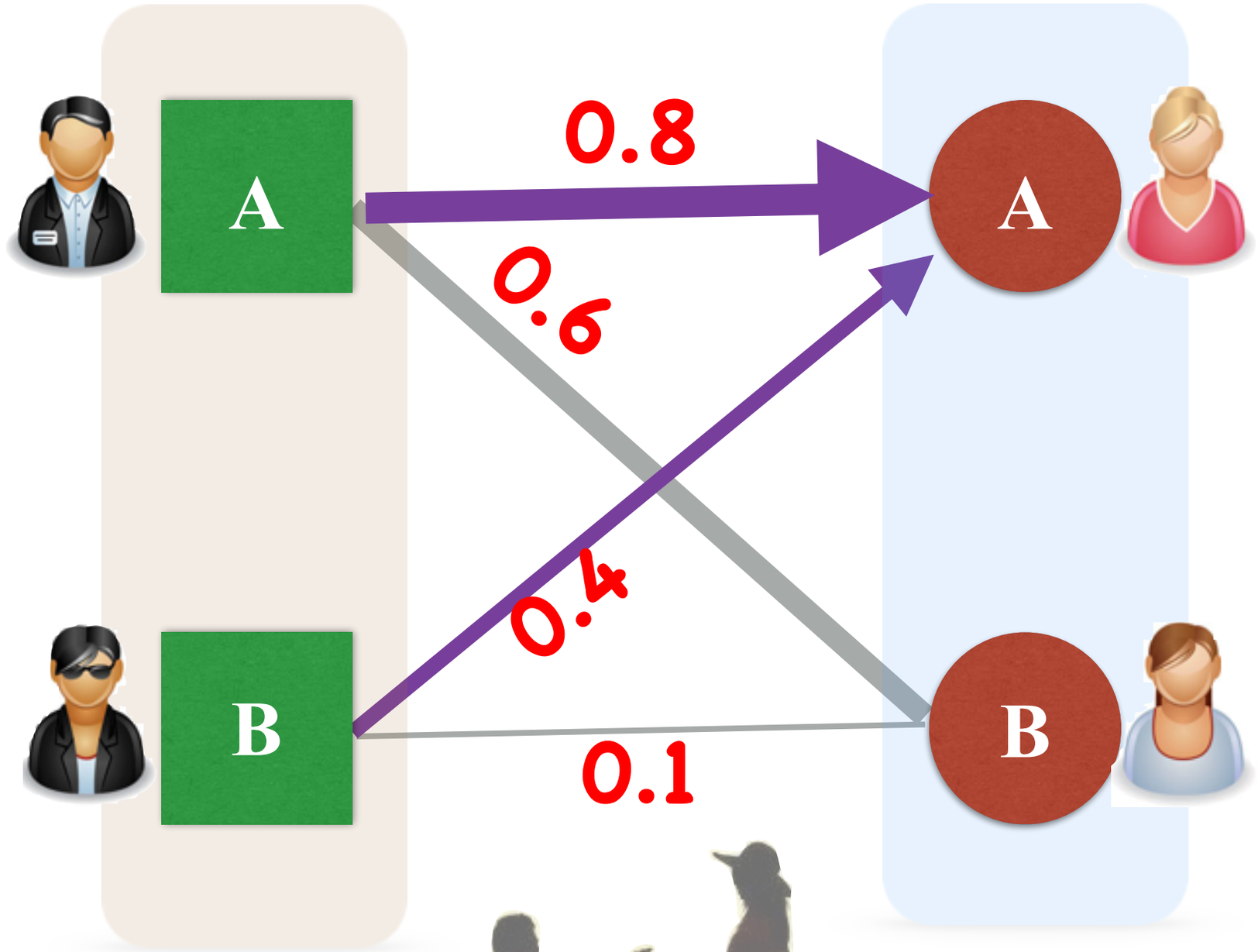


# A Blocking Pair is Unstable

# Stable Matching/Marriage

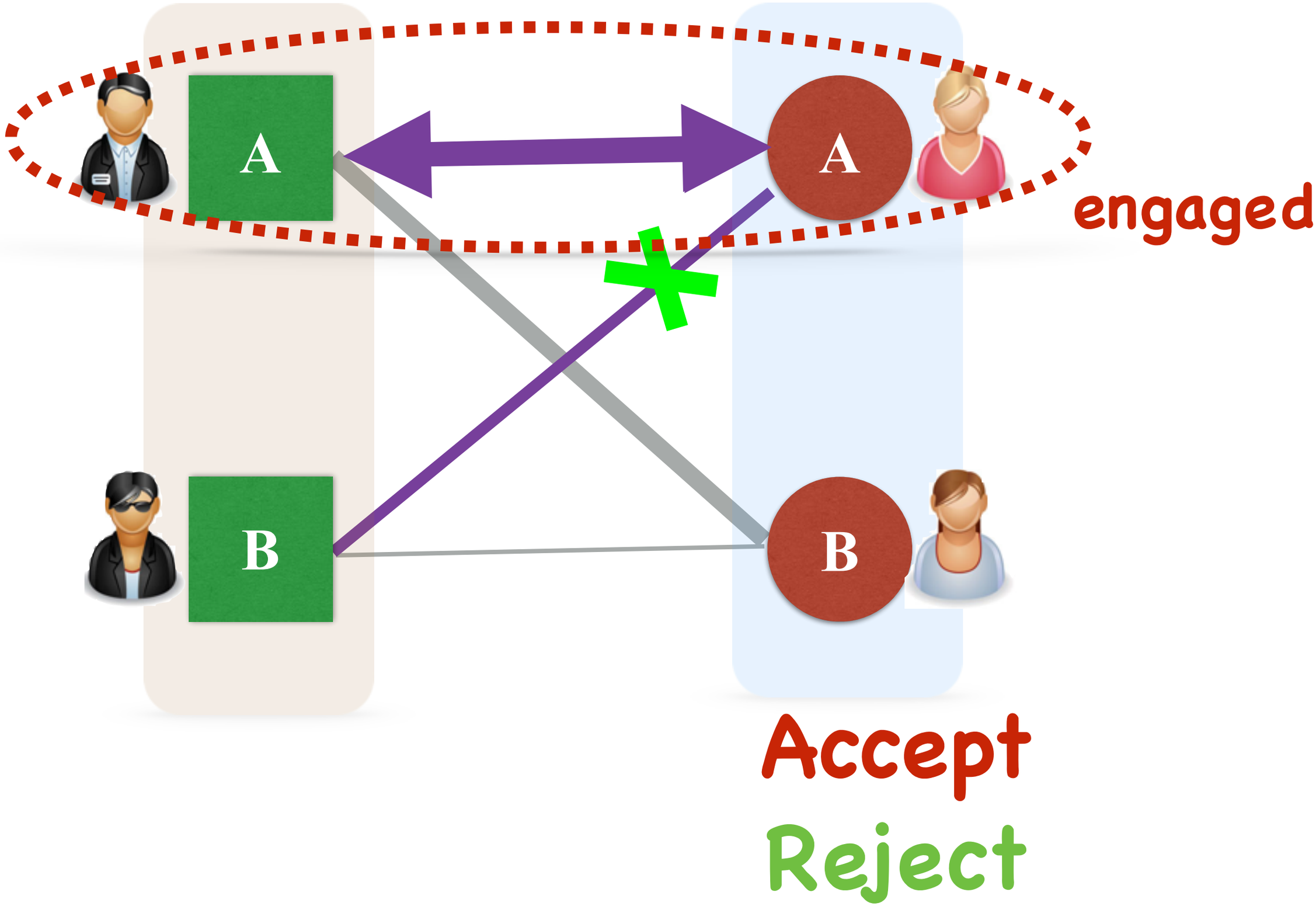


# Stable Matching Algorithm

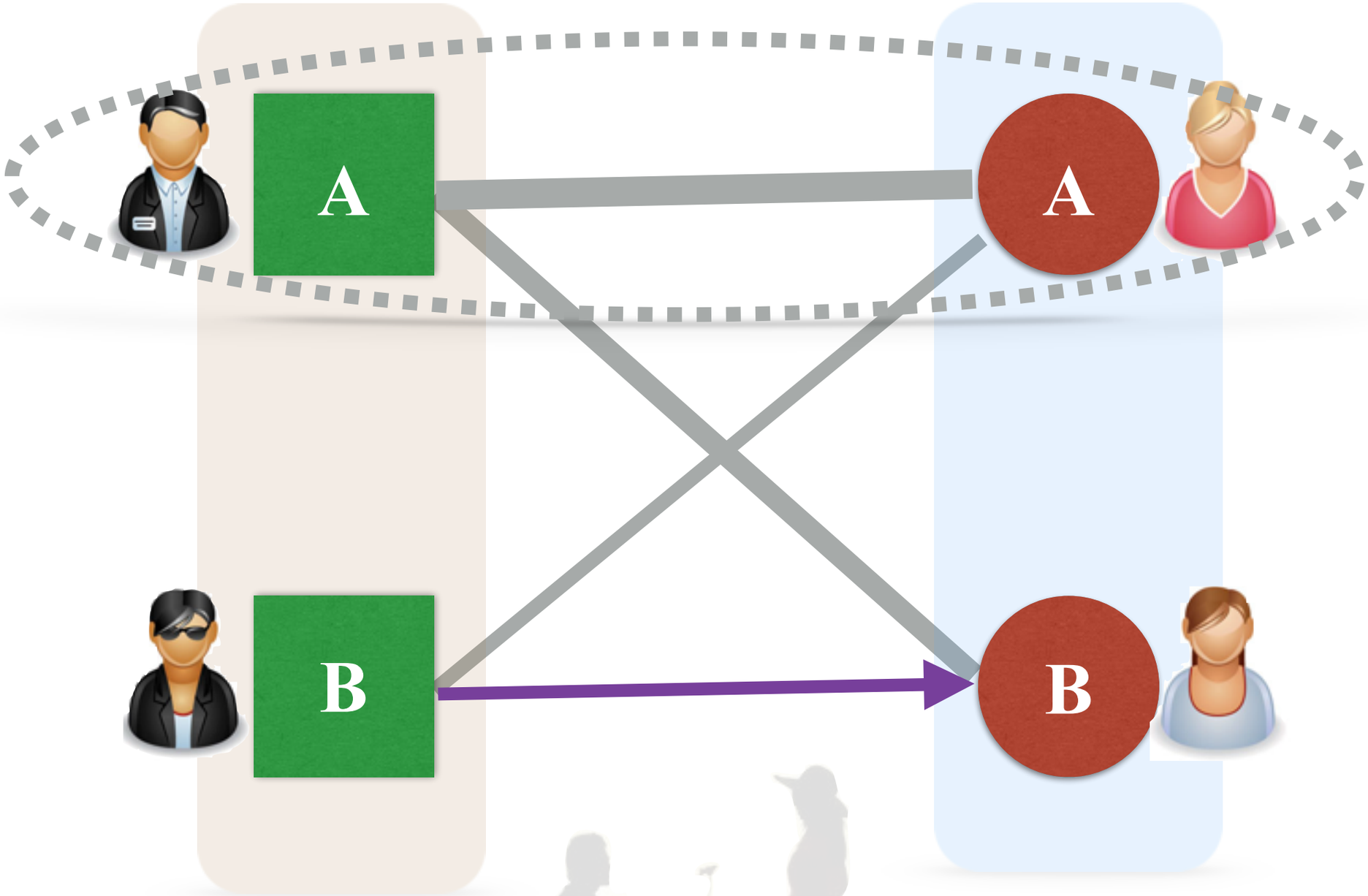



Propose

# Stable Matching Algorithm

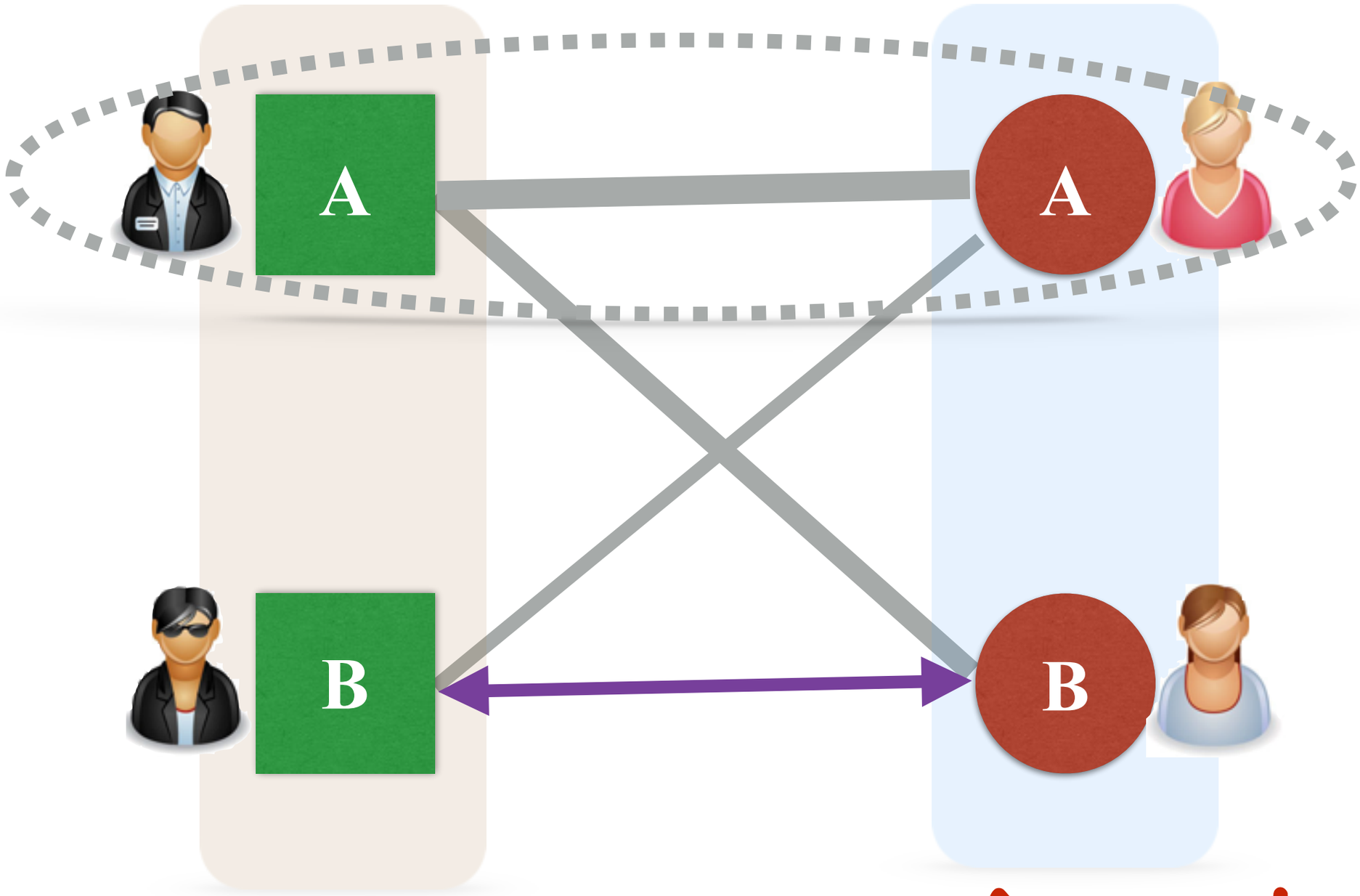


# Stable Matching Algorithm




Propose

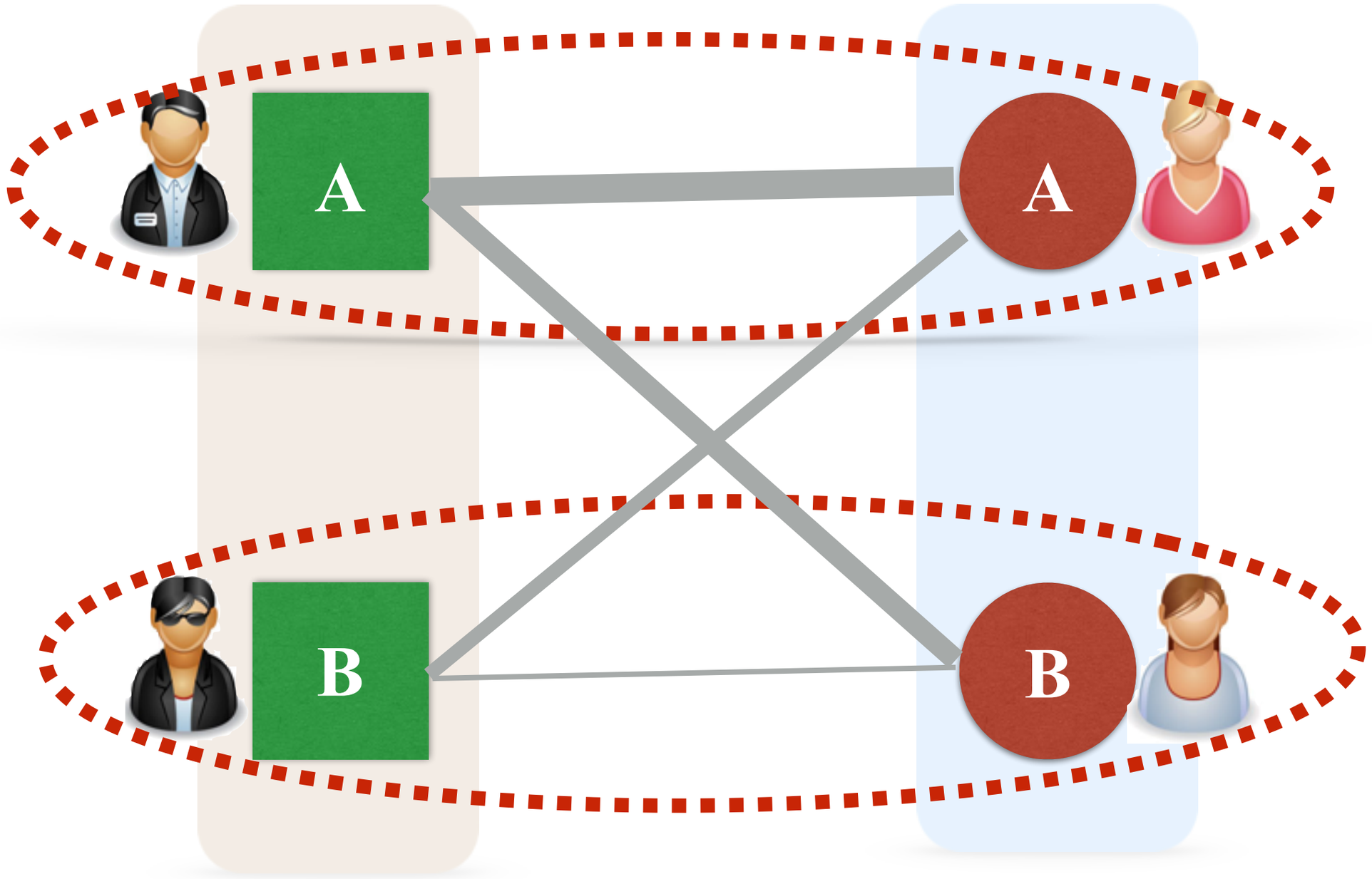
# Stable Matching Algorithm




Accept

Reject

# Stable Matching Algorithm

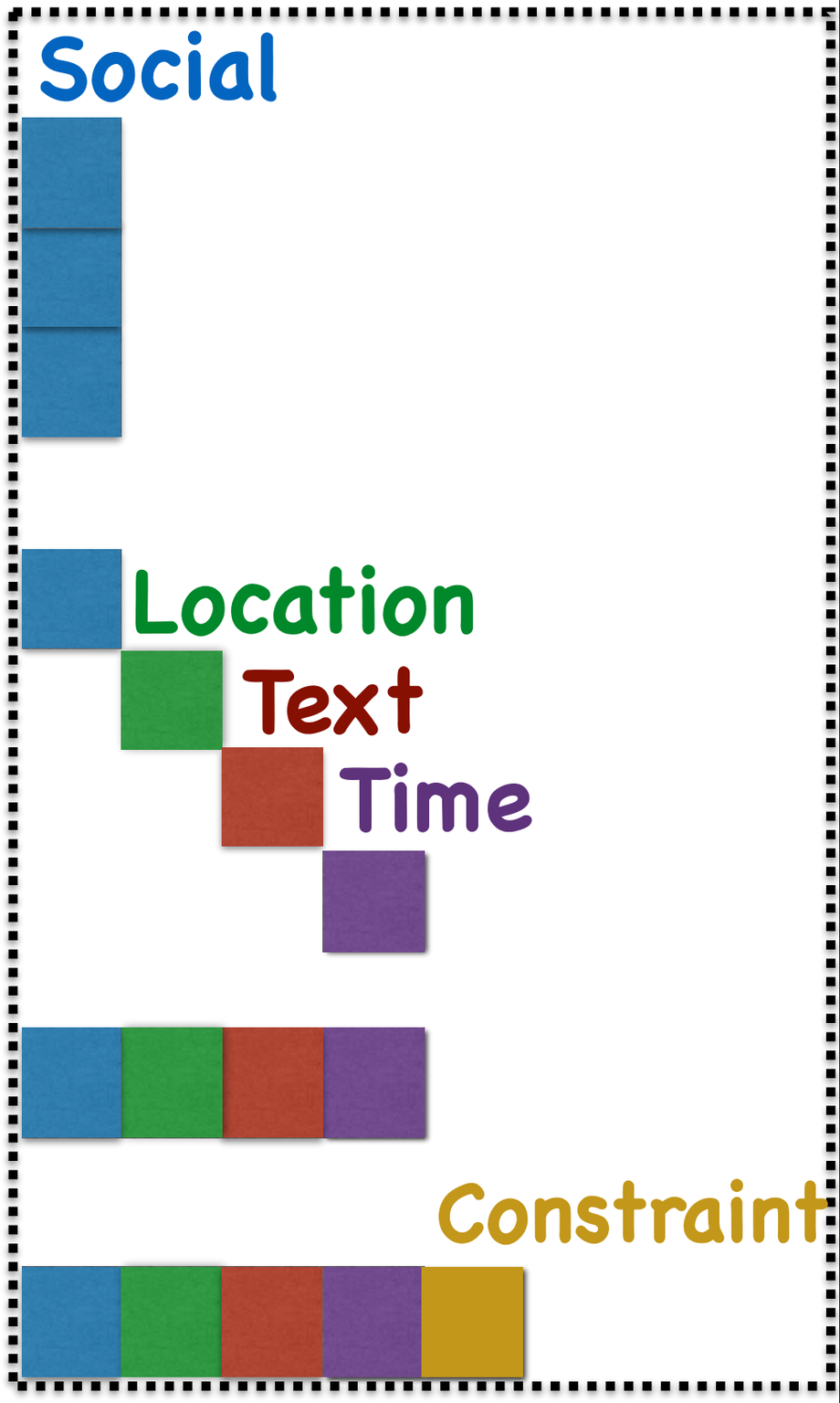


# Data Sets

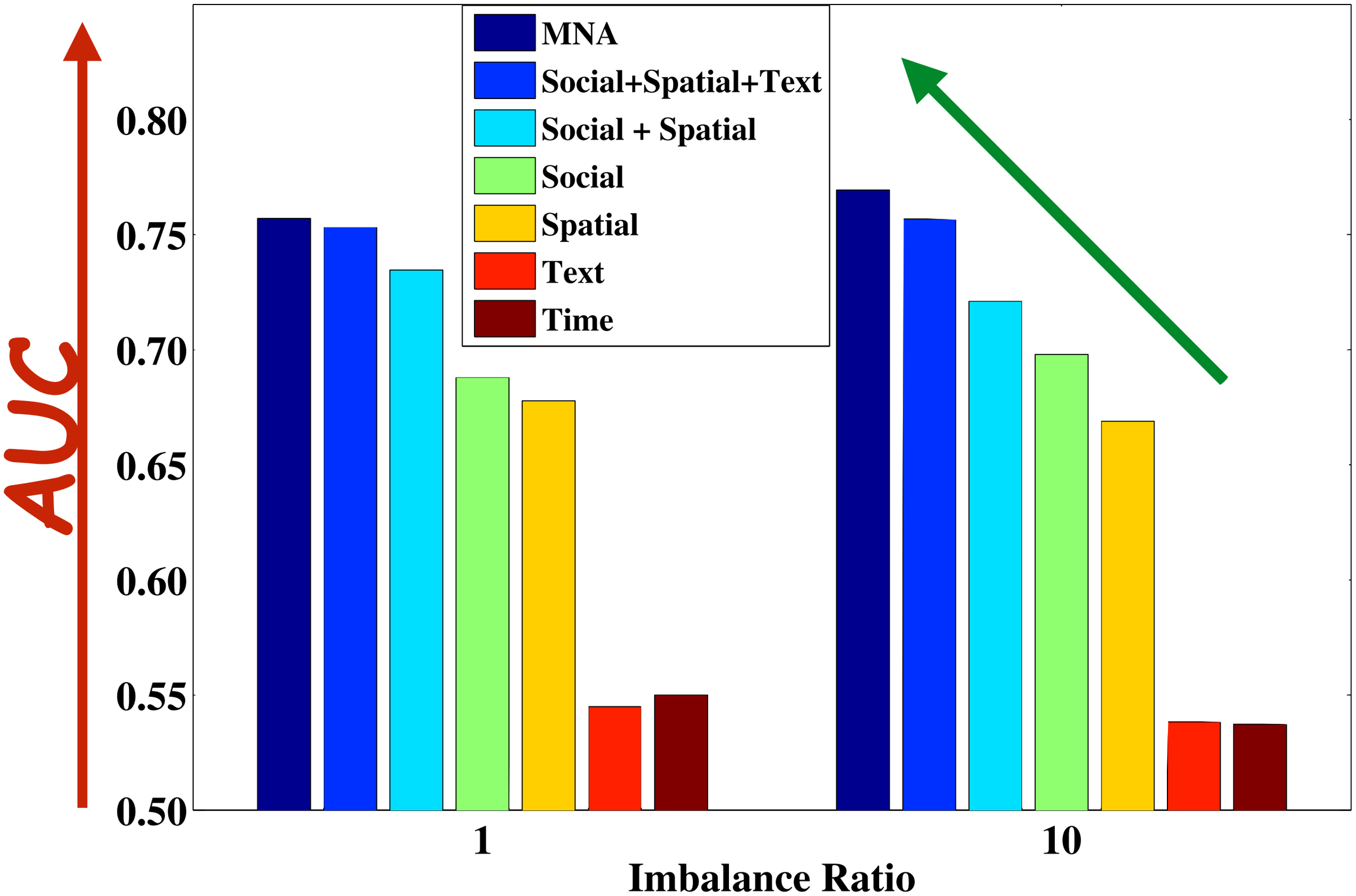
| | | network | |
|---|---|---|---|
| | property | **Twitter** | **Foursquare** |
| # node | user | 5,223 | 5,392 |
| | tweet/tip | 9,490,707 | 48,756 |
| | location | 297,182 | 38,921 |
| # link | friend/follow | 164,920 | 31,312 |
| | write | 9,490,707 | 48,756 |
| | locate | 615,515 | 48,756 |

# Compared Methods

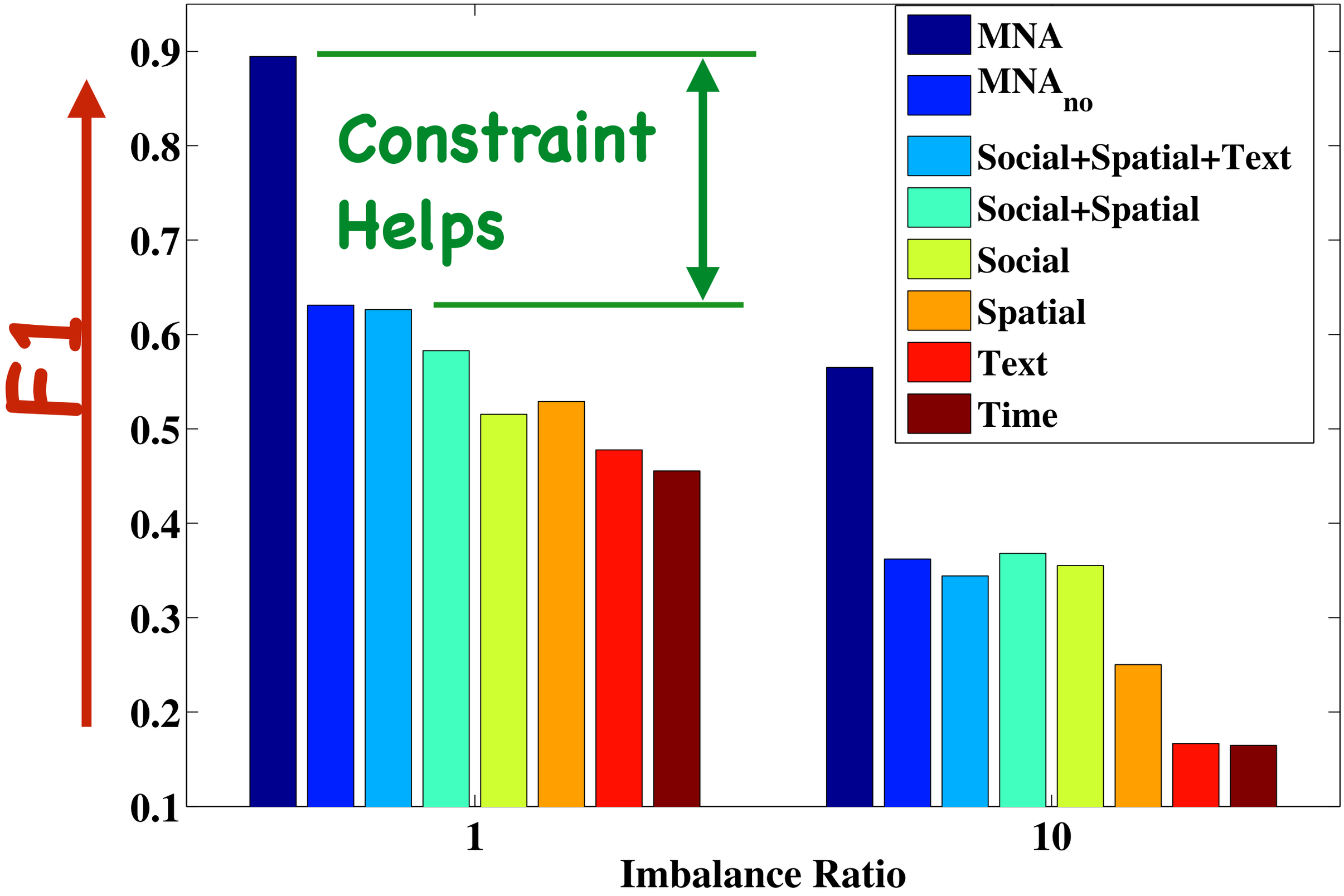
- **Unsupervised Link Prediction**
  - CN: Common Neighbor
  - JC: Jaccard Coefficient
  - AA: Adamic/Adar
- **Supervised Link Prediction**
  - Social Features
  - Spatial Features
  - Text Features
  - Temporal Features
- Multi-Network Anchoring w/o Constraint
- **Multi-Network Anchoring**



# More Features are better



# Constraint is better



# #labeled data

| measure | | number of labeled anchor links | | | | |
|---|---|---|---|---|---|---|
| | methods | 10 | 20 | 30 | 40 | 50 |
| Auc | Mna | **0.556±0.029** | **0.640±0.040** | **0.657±0.021** | **0.688±0.021** | **0.705±0.008** |
| | Social | 0.507±0.015 | 0.534±0.021 | 0.572±0.029 | 0.628±0.029 | 0.627±0.039 |
| | Spatial | 0.549±0.061 | 0.621±0.046 | 0.602±0.043 | 0.658±0.005 | 0.651±0.017 |
| | Text | 0.529±0.005 | 0.533±0.031 | 0.510±0.043 | 0.530±0.003 | 0.544±0.006 |
| | Time | 0.538±0.006 | 0.539±0.011 | 0.534±0.017 | 0.519±0.024 | 0.543±0.006 |
| | CN | 0.527±0.005 | 0.541±0.004 | 0.581±0.007 | 0.591±0.003 | 0.599±0.004 |
| | JC | 0.528±0.007 | 0.546±0.004 | 0.577±0.010 | 0.593±0.007 | 0.608±0.010 |
| | AA | 0.524±0.004 | 0.552±0.008 | 0.575±0.007 | 0.585±0.012 | 0.601±0.010 |
| F1 | Mna | **0.735±0.055** | **0.828±0.035** | **0.843±0.036** | **0.849±0.027** | **0.862±0.012** |
| | Mna_no | 0.502±0.083 | 0.510±0.095 | 0.522±0.032 | 0.584±0.021 | 0.584±0.042 |
| | Social | 0.031±0.063 | 0.190±0.110 | 0.334±0.044 | 0.382±0.030 | 0.396±0.026 |
| | Spatial | 0.259±0.317 | 0.430±0.197 | 0.455±0.267 | 0.425±0.203 | 0.592±0.161 |
| | Text | 0.466±0.018 | 0.493±0.038 | 0.457±0.057 | 0.490±0.057 | 0.435±0.018 |
| | Time | 0.559±0.011 | 0.553±0.021 | 0.529±0.036 | 0.485±0.080 | 0.523±0.061 |

# Imbalance Ratio

| measure | methods | 1 | 2 | 3 |
|---|---|---|---|---|
| Auc | Mna | **0.757±0.010** | **0.771±0.008** | **0.751±0.011** |
| | Social | 0.688±0.061 | 0.680±0.046 | 0.711±0.025 |
| | Spatial | 0.678±0.012 | 0.659±0.011 | 0.666±0.002 |
| | Text | 0.545±0.012 | 0.546±0.005 | 0.542±0.004 |
| | Time | 0.550±0.006 | 0.542±0.008 | 0.530±0.012 |
| | CN | 0.656±0.014 | 0.638±0.008 | 0.634±0.009 |
| | JC | 0.665±0.007 | 0.661±0.004 | 0.651±0.008 |
| | AA | 0.641±0.004 | 0.649±0.004 | 0.654±0.007 |
| F1 | Mna | **0.895±0.008** | **0.839±0.015** | **0.751±0.014** |
| | Mna_no | 0.631±0.014 | 0.584±0.006 | 0.525±0.009 |
| | Social | 0.515±0.026 | 0.485±0.015 | 0.474±0.016 |
| | Spatial | 0.529±0.179 | 0.492±0.100 | 0.394±0.086 |
| | Text | 0.478±0.050 | 0.385±0.013 | 0.337±0.018 |
| | Time | 0.455±0.045 | 0.380±0.011 | 0.353±0.028 |

# Summary

Predicting Anchor Links across Multiple Heterogeneous Social Networks

- extract heterogeneous features: social, spatial, temporal

- predicting anchor links: stable matching

# Q&A