Organizational Chart Inference

Jiawei Zhang[1], Philip S. Yu[1,2], and Yuanhua Lv[3]

[1] University of Illinois at Chicago, IL, USA
[2] Institute for Data Science, Tsinghua University, Beijing, China
[3] Microsoft Research, Redmond, WA, USA

People in Organizations are Hierarchized into Tree-Structured **Organizational Charts**

Company: AMD

Enterprise Social Networks (ESNs) are Launched in Many Companies (An example: Yammer)

- Over 8 million registered user
- More than 200,000 companies

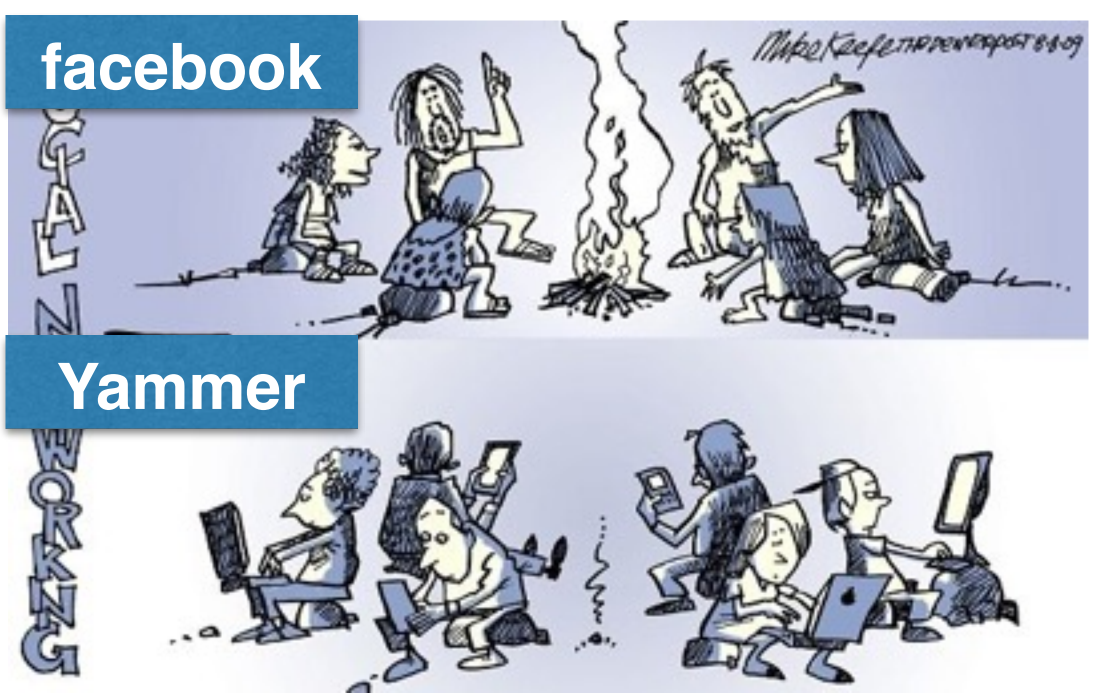
Enterprise Social Networks(ESNs) are **New** and **Different**

What are Enterprise Social Networks?

- A kind of online social networks that provide employees with various integrated professional services to help deal with daily work issues.

Functions of Enterprise Social Networks

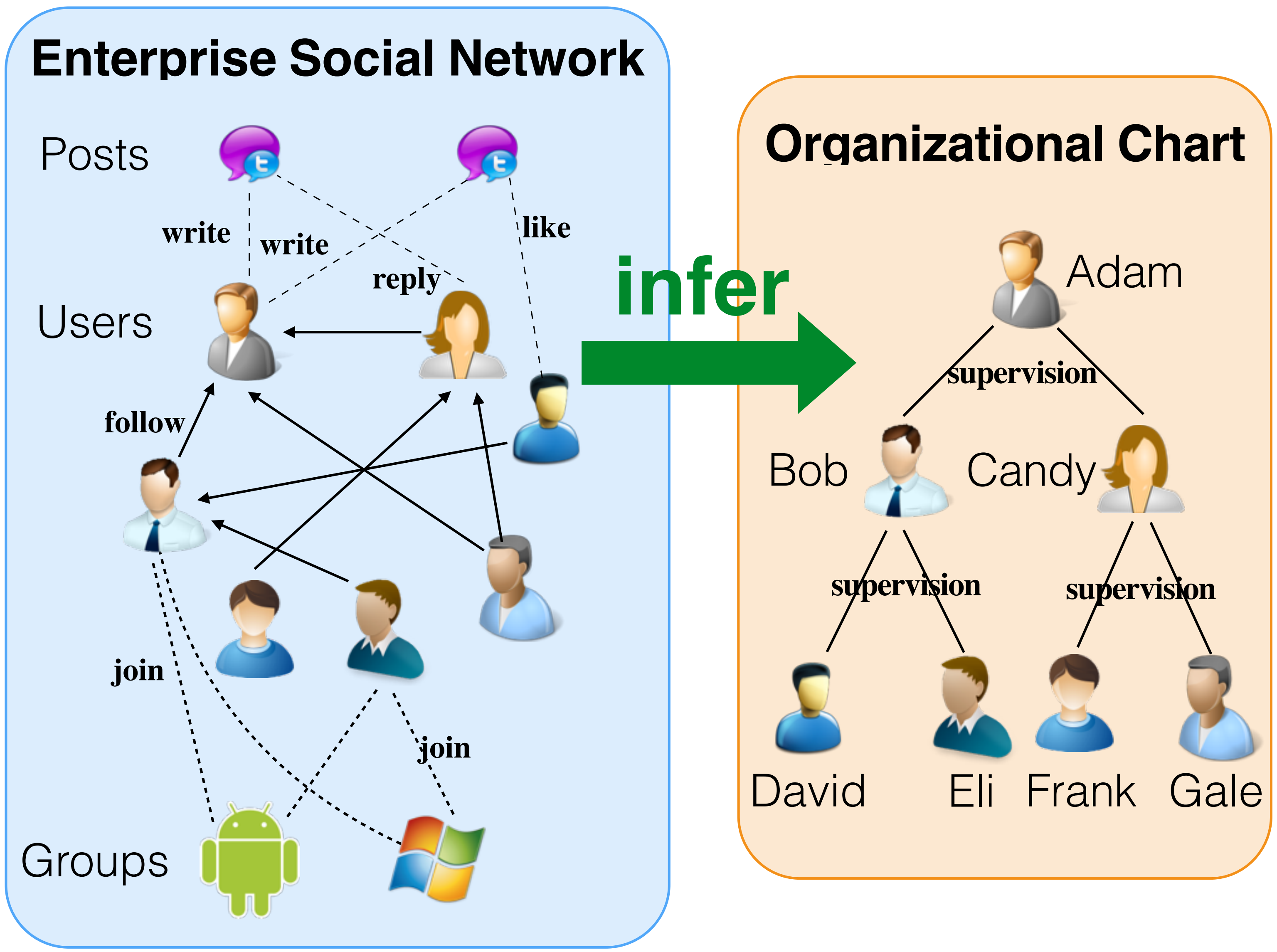
- better project management
- easier communication among employees
- broader information sharing
- more effective cooperator identification
- …



Different from Traditional Social Networks:

- Facebook: Casual and Personal
- Yammer: Formal and Professional

Problem Studied: **Organizational Chart Inference**



Why Study this Problem:

- Complete organizational chart of most companies are confidential due to privacy and security concerns.

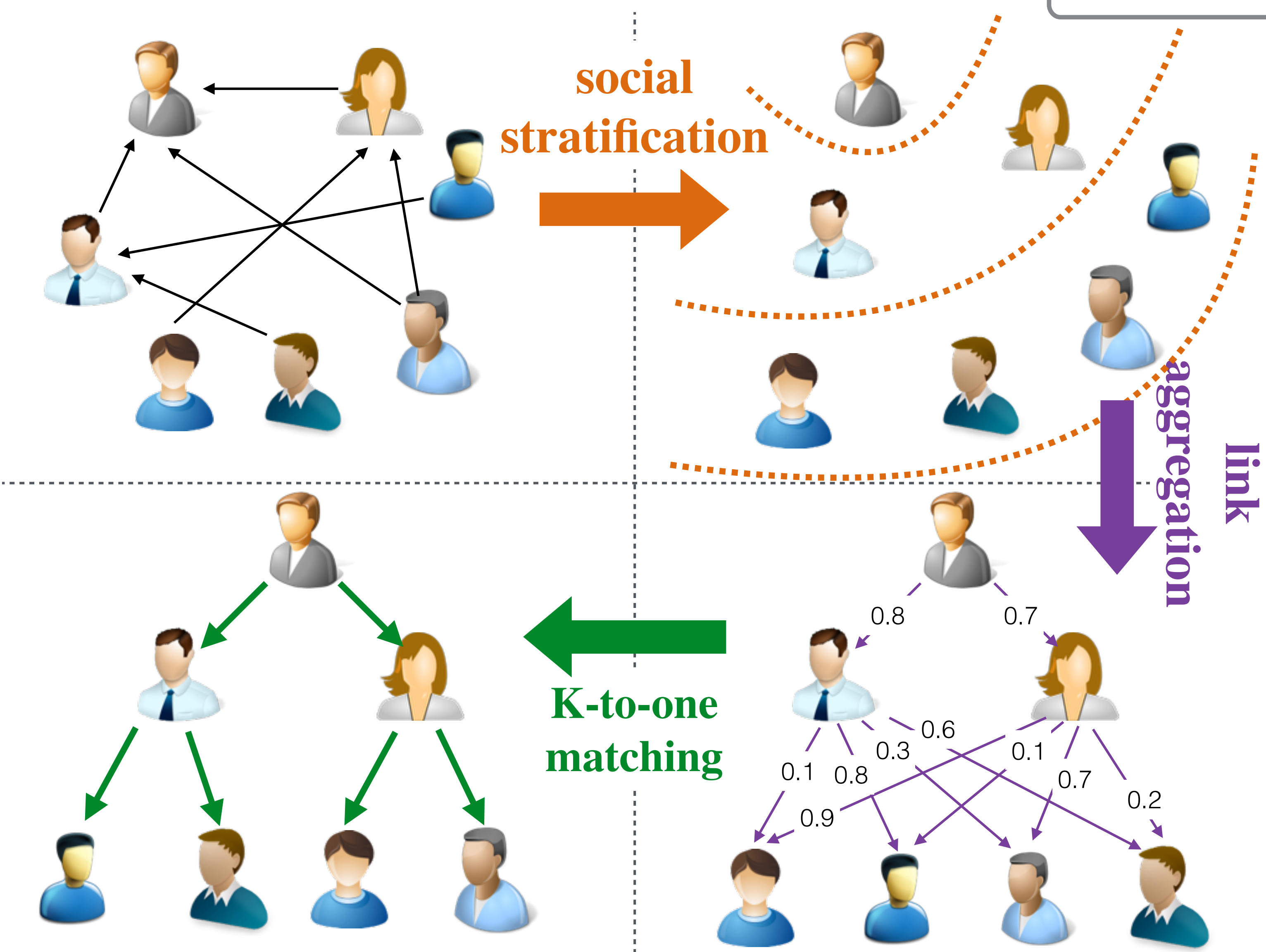
Other Concrete Applications:

- identifying command structures of terrorist organizations based on communication/traffic networks of their members.
- inferring the social hierarchies of animals based on their observed interaction networks.

Proposed Framework: **CREATE** (ChArT REcovEr)

step 1: partition users into different hierarchical tiers

Motivation: supervision links only exist from manager to subordinates, pre-stratification can identify their relative management relationships and shrink the inference space greatly



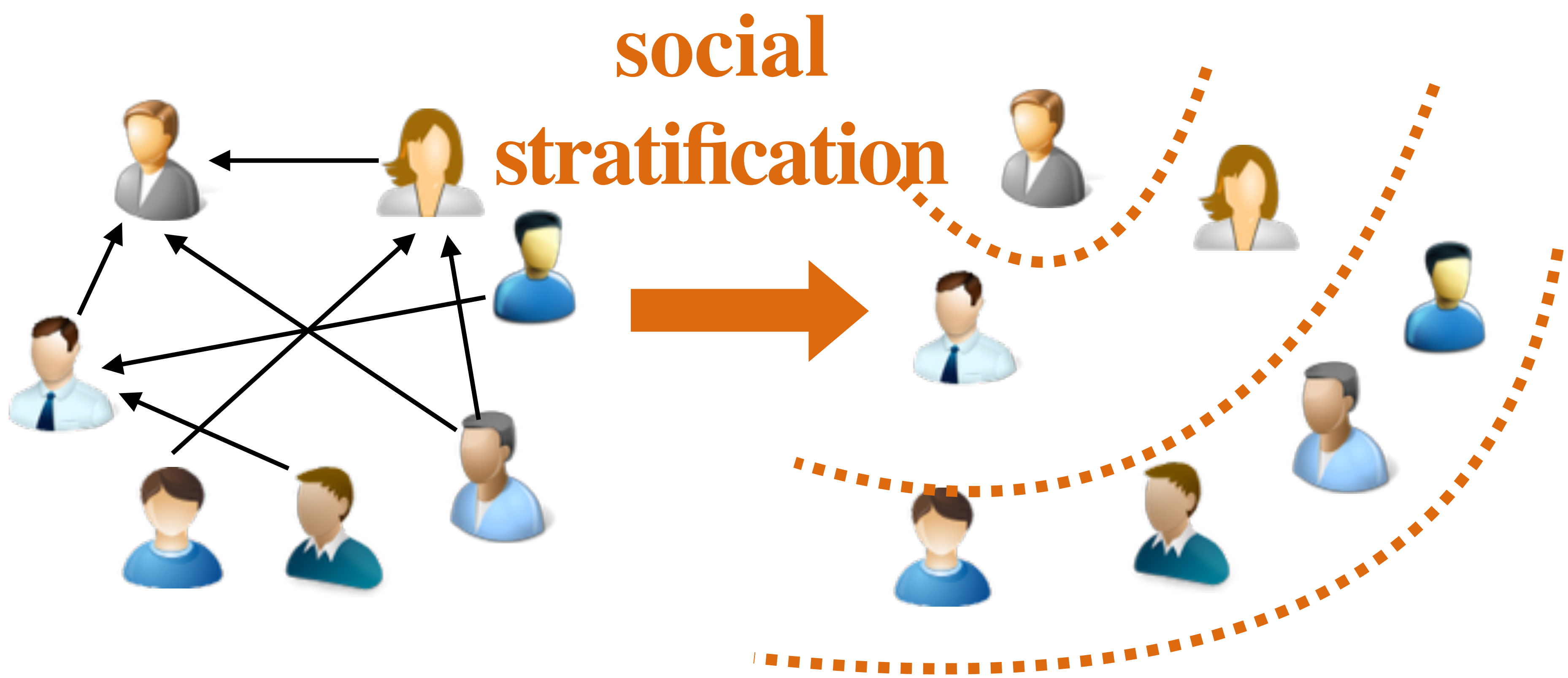
step 2: supervision link inference with heterogeneous information in enterprise social networks

Motivation: heterogeneous information in online ESNs can capture potential management relationships in various aspects. How to use it is challenging.

step 3: K-to-1 matching to prune redundant supervision links

Motivation: (1) the number of subordinates each manager can supervise is limited; (2) each subordinate has one single direct manager

Challenge 1: How to Stratify Users in ESNs



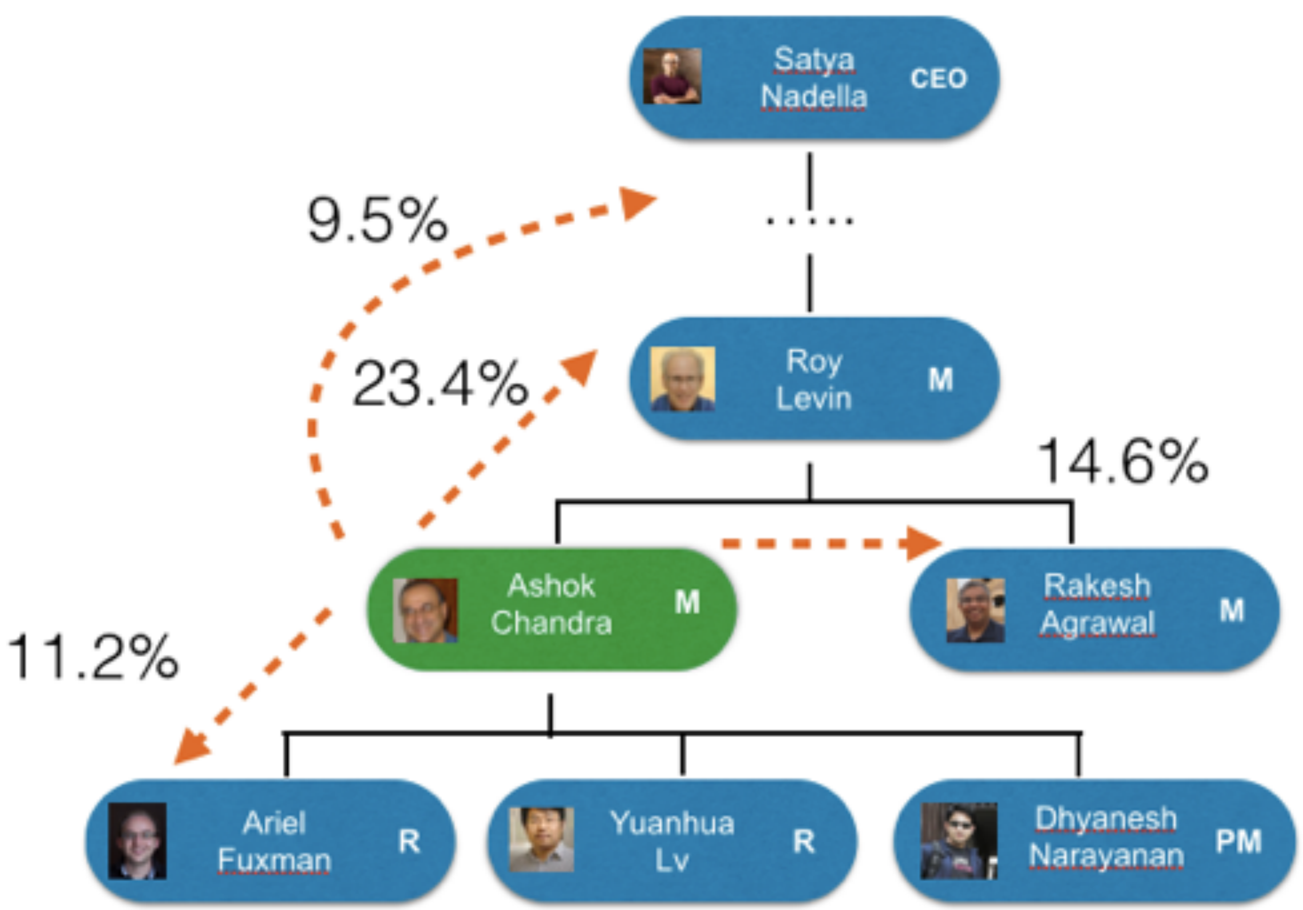
- **Definition** (*Social Class*): In this paper, we define social class of users in online ESNs as their management level in the company, where CEO belongs to social class 1, EVPs belong to class 2, and so forth.

Social Stratification aims at inferring the mapping $c : \mathcal{U} \rightarrow \mathbb{Z}^+$ to project users in ESNs to their social classes. Social class of user u can be represented as

$$c(u) = \begin{cases} 1, & \text{if } u \text{ is the CEO;} \\ c\left(m(u)\right) + 1, & \text{otherwise.} \end{cases}$$

Proposed Method: utilize social connections among users in ESNs

- Observation 1: Users like to follow their managers
  - **Definition** (*Class Transcendence Social Link*): Social link (u, v) (u follows v) is defined as a class transcendence social link in online ESNs iff c(u) < c(v).



- Class transcendence penalty of social link (u, v) can be represented as

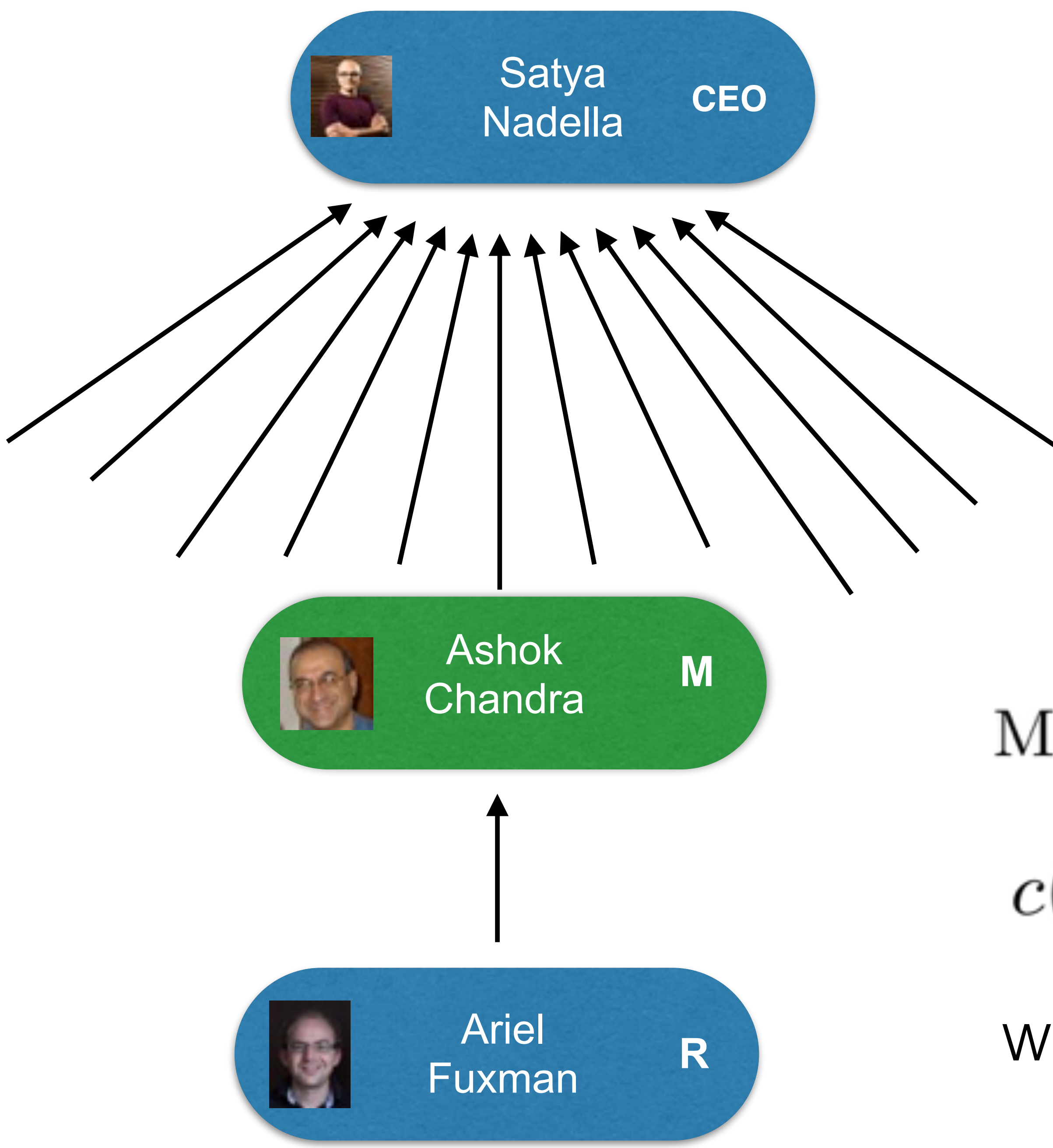
$$p\left(c(u), c(v)\right) = \begin{cases} 0, & \text{if } c(u) > c(v) \\ c(v) - c(u) + 1, & \text{otherwise.} \end{cases}$$

- Class transcendence penalty introduced by all social link in the network can be represented as

$$p\left(c(\mathcal{U})\right) = \sum_{(u,v)\in\mathcal{S}} p\left(c(u), c(v)\right) = \sum_{(u,v)\in\mathcal{S}} \max\{c(v) - c(u) + 1, 0\}.$$

Proposed Method: utilize social connections among users in ESNs

- Observation 2: Matthew Effect ("The rich get richer") - people at higher management level can have more followers



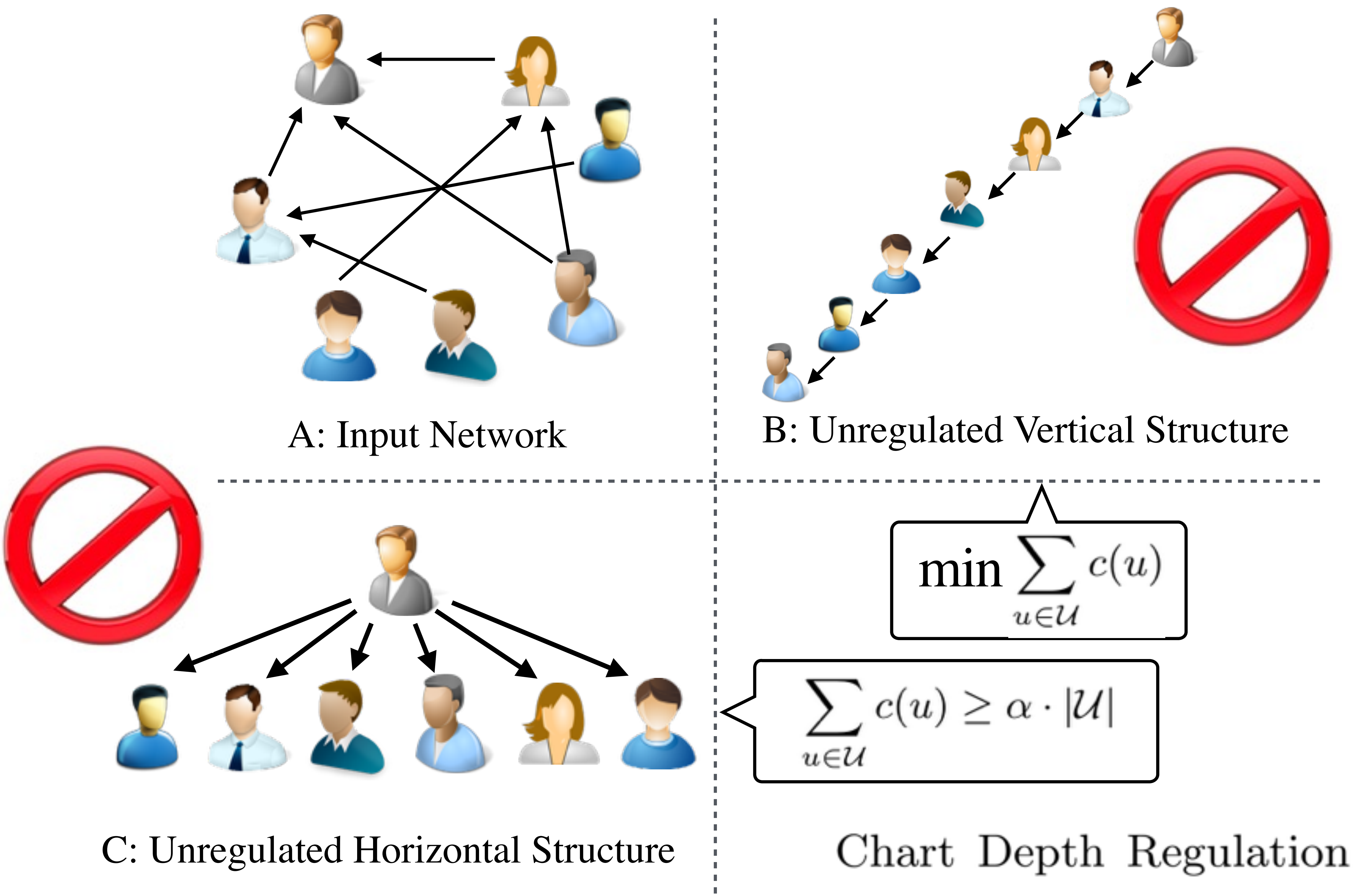
Matthew Effect based Constraint

$$c(u) \leq c(v) \text{ if } |\Gamma(u)| \geq |\Gamma(v)|$$

where $\Gamma(u)$ denotes the followers of u

Proposed Method: utilize social connections among users in ESNs

- Observation 3: Organizational Chart depth need regulation



Social Stratification Objective Function

the optimal mapping $c : \mathcal{U} \rightarrow \mathbb{Z}^+$ can be represented as

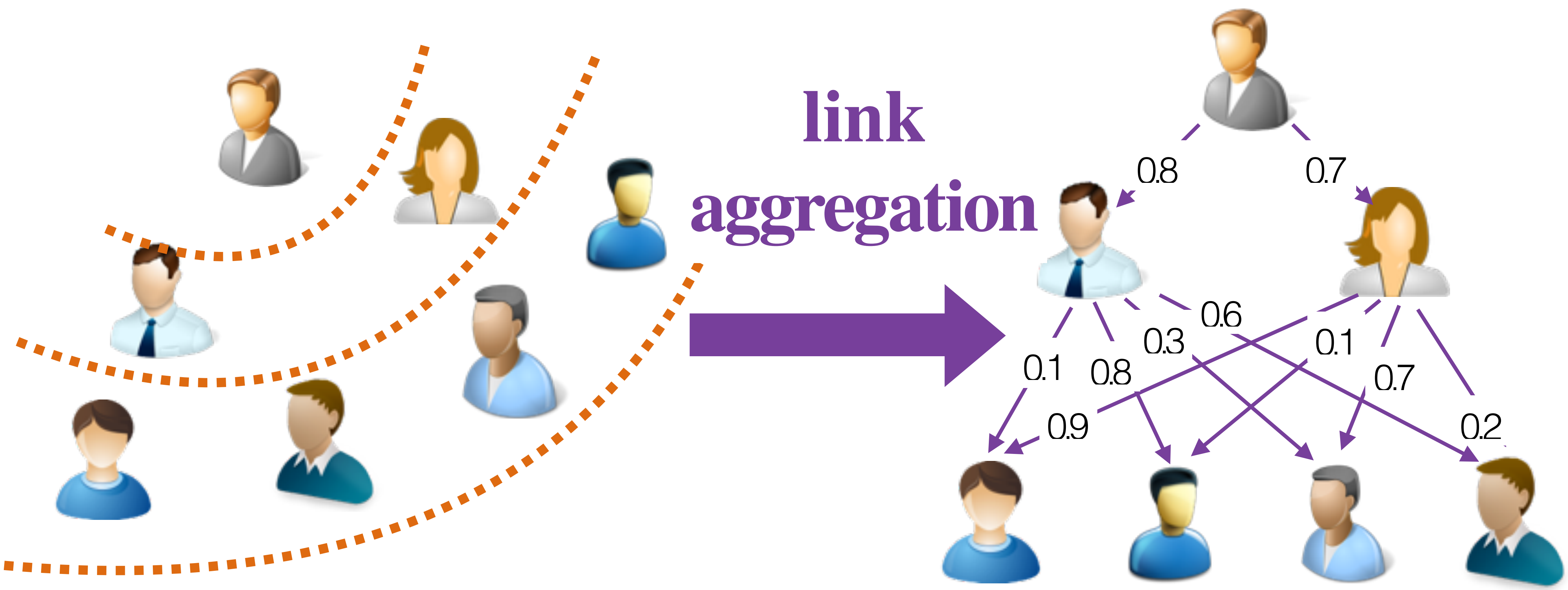
$$c^*(\mathcal{U}) = \arg \min_{\{c(u_1), c(u_2), \cdots, c(u_{|\mathcal{U}|})\}} \sum_{(u,v)\in\mathcal{S}} p\left(c(u), c(v)\right) + \sum_{u\in\mathcal{U}} c(u),$$

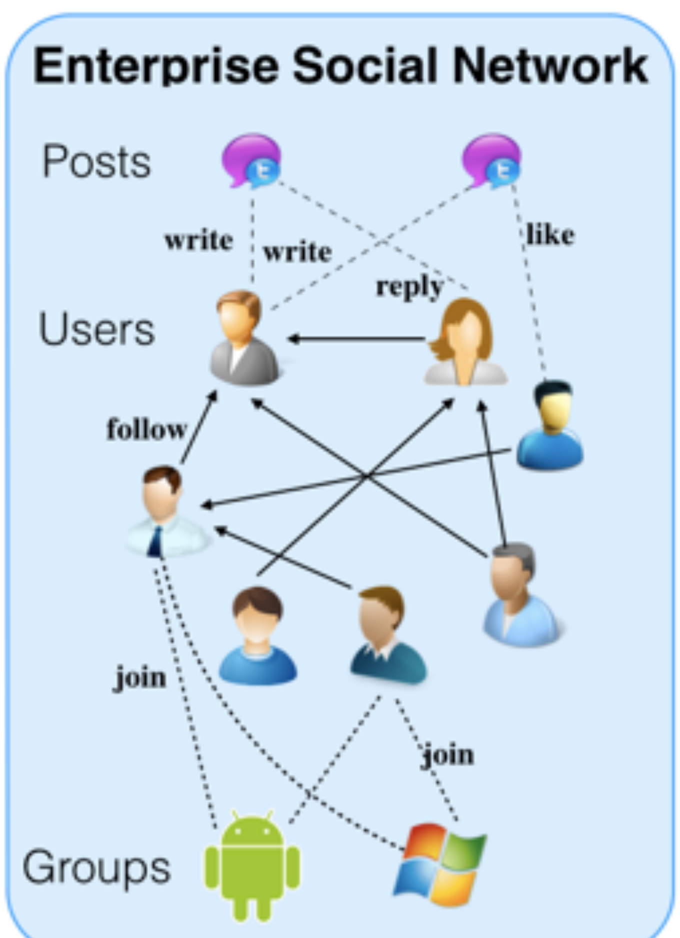
$$\begin{aligned}
s.t., \; & p(c(u), c(v)) \geq c(v) - c(u) + 1, \forall (u,v) \in \mathcal{S}, \\
& p(c(u), c(v)) \geq 0, \forall (u,v) \in \mathcal{S}, \\
& c(u) \leq c(v), \forall u, v \in \mathcal{U}, \text{ if } |\Gamma(u)| \geq |\Gamma(v)|, \\
& \sum_{u\in\mathcal{U}} c(u) \geq \alpha \cdot |\mathcal{U}|, \\
& c(u) = 1, \text{ if } u \text{ is the CEO}, \\
& c(u) > 1, c(u) \in \mathbb{Z}^+, \forall u \in \mathcal{U} \setminus \{\text{CEO}\}, \\
& p\left(c(u), c(v)\right) \in \mathbb{Z}, \forall (u,v) \in \mathcal{S}.
\end{aligned}$$

Class Transcendence Penalty

Definitions of “Social Class” and “Penalty”

Challenge 2: Supervision Link Inference with **Heterogeneous Information**





Heterogeneous Information exists in ESNs

- Social Links: users connected by social links are more likely to be colleagues.
- Group Membership: colleagues may have similar expertise and can join common technique groups.
- Online Communications: colleagues may communicate more often online via writing posts, replies, like posts published by their colleagues.

Proposed Method: Meta Paths Aggregation

Meta Paths based on heterogeneous information in online ESNs

- *Follow*: User $\xrightarrow{follow}$ User, whose notation is "$U \to U$" or $\Phi_1(U, U)$.
- *Follower of Follower*: User $\xrightarrow{follow}$ User $\xrightarrow{follow}$ User, whose notation is "$U \to U \to U$" or $\Phi_2(U, U)$.
- *Common Followee*: User $\xrightarrow{follow}$ User $\xrightarrow{follow^{-1}}$ User, whose notation is "$U \to U \leftarrow U$" or $\Phi_3(U, U)$.
- *Common Follower*: User $\xrightarrow{follow^{-1}}$ User $\xrightarrow{follow}$ User, whose notation is "$U \leftarrow U \to U$" or $\Phi_4(U, U)$.
- *Common Group Membership*: User $\xrightarrow{join}$ Group $\xrightarrow{join^{-1}}$ User, whose notation is "$U \to G \leftarrow U$" or $\Phi_5(U, U)$.
- *Reply Post*: User $\xrightarrow{write}$ Post $\xrightarrow{reply}$ Post $\xrightarrow{write^{-1}}$ User, whose notation is "$U \to P \to P \leftarrow U$" or $\Phi_6(U, U)$.
- *Like Post*: User $\xrightarrow{write}$ Post $\xrightarrow{like^{-1}}$ User, whose notation is "$U \to P \to P \leftarrow U$" or $\Phi_7(U, U)$.

- confidence of potential supervision link (u, v) based on the i_{th} meta path can be represented as

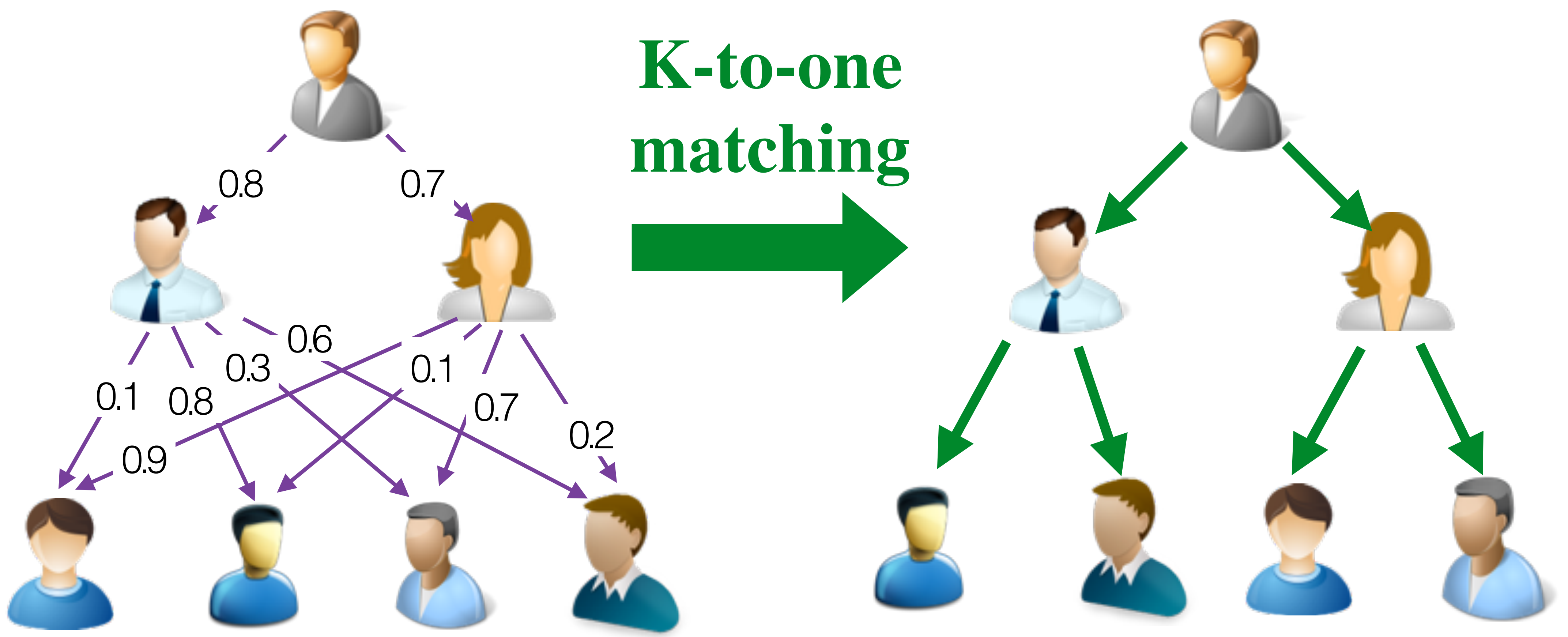
$$\text{DP-intimacy}_i(u,v) = \frac{|\mathcal{PATH}_i(u \rightsquigarrow v)| + |\mathcal{PATH}_i(v \rightsquigarrow u)|}{|\mathcal{PATH}_i(u \rightsquigarrow \cdot)| + |\mathcal{PATH}_i(v \rightsquigarrow \cdot)|}$$

- overall intimacy between users u and v can be represented as

$$intimacy(u,v) = \frac{e^{\sum_{(i)} \omega_i \text{DP-intimacy}_i(u,v)}}{1 + e^{\sum_{(i)} \omega_i \text{DP-intimacy}_i(u,v)}} \in [0,1],$$

where the value of ω_i denotes the weight of social meta path Φ_i and $\sum_i \omega_i = 1$.

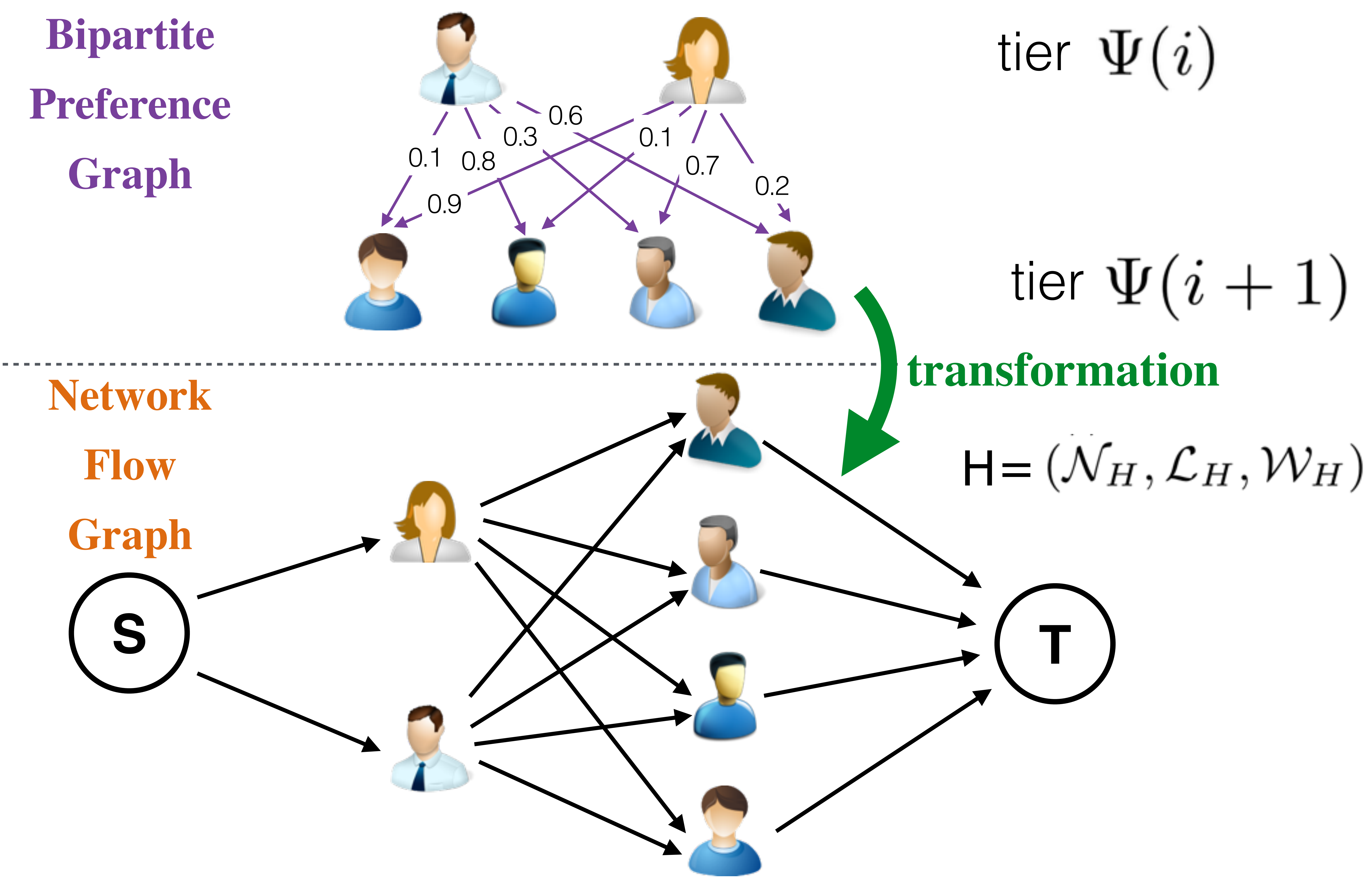
Challenge 3: Regulated Supervision Workload Allocation



Supervision Workload Allocation

- each supervisor can manager limited number of subordinates, whose number is K
- each subordinate on needs to report to one manager
- As a result, the constraint on supervision links is K-to-1

Proposed Method: Network Flow based Consecutive Social Class Matching



Minimum Cost Network Flow based Social Class Matching Objective Function

$$\min \sum_{(u,v)\in\Lambda(i,i+1)} x_{u,v}(1 - intimacy(u,v))$$

$$s.t.\; 0 \le x_{s,u} \le K, \text{for } \forall u \in \Psi(i),$$

$$x_{v,t} = 1, \text{for } \forall v \in \Psi(i+1),$$

$$x_{u,v} \in \{0,1\}, \text{for } \forall u \in \Psi(i), \forall v \in \Psi(i+1),$$

$$\sum_{w\in\mathcal{N}_H,(w,u)\in\mathcal{L}_H} x_{w,u} = \sum_{v\in\mathcal{N}_H,(u,v)\in\mathcal{L}_H} x_{u,v}, \forall u \in \mathcal{N}_H.$$

minimize the cost (

K-to-1 co

Mass Balance Constraint:
input network flow
=
output network flow

Experiments

- Dataset[1]
  - Yammer used in Microsoft
    - covers all the user-generated content (such as posts, replies, topics, etc.) and social graphs (such as user-user following links, user-group memberships, user-topic following links, etc.)
  - Complete organizational chart of all employees in Microsoft
    - includes more than 100k Microsoft employees
  - All the users in yammer are registered with the official employment ID in Microsoft, via which we can identify them in the organizational chart correspondingly.

[1] We are not able to reveal the actual numbers here and throughout the paper for commercial reasons.

Experiment Settings

- Ground truth
  - the complete organization chart is used for evaluation only
  - unsupervised learning setting
- Step 1: Social Stratification Comparison Methods
  - CREATE: proposed in this paper
  - ASD (Agony based Social Division): proposed to detect the social hierarchies of regular users in general online social networks
- Evaluation Metrics
  - Accuracy, Precision, Recall,
  - MAE (Mean Absolute Error), MSE (Mean Square Error), and R^2

Experiment Settings

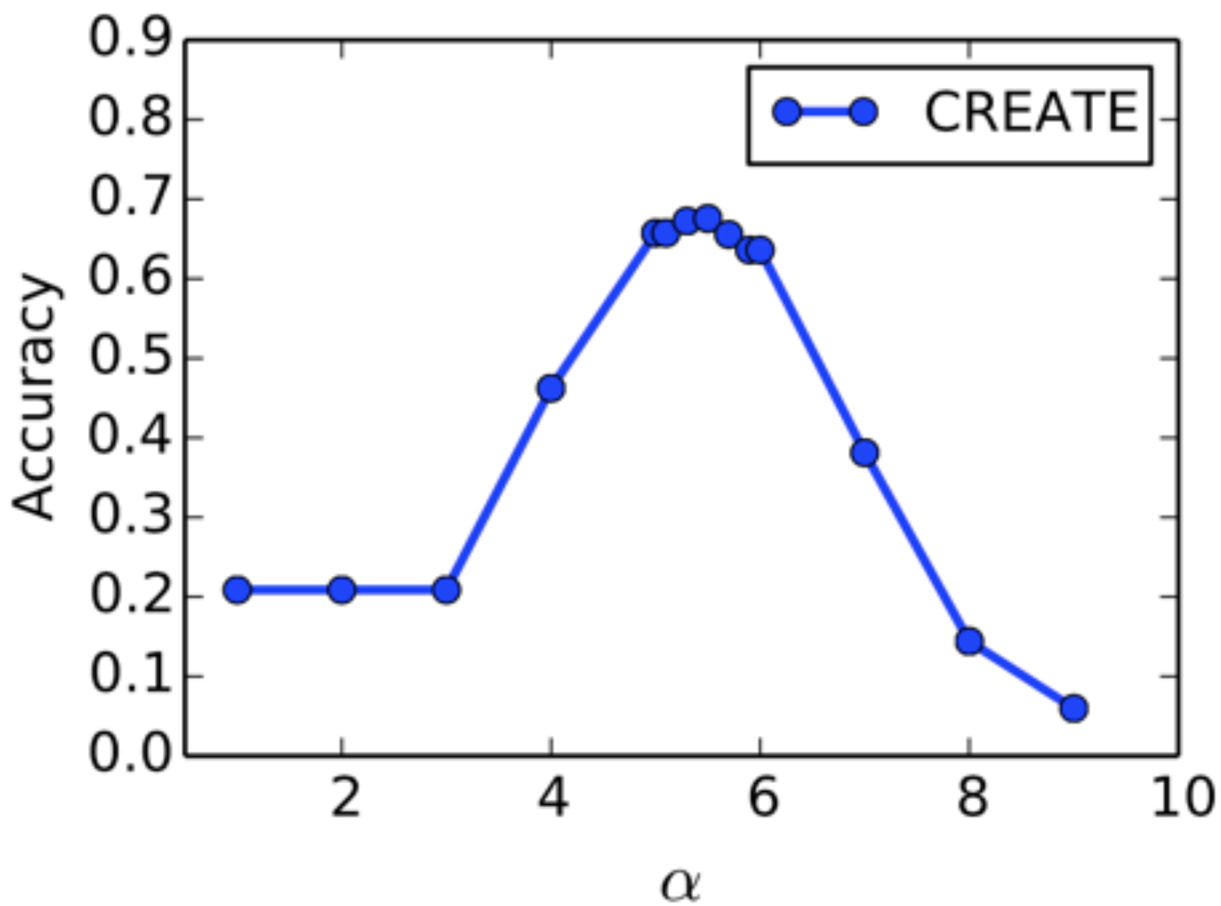
- Organizational Chart Inference Comparison Methods (No Existing Methods can Solve the Problem)
  - CREATE: consists of 3 steps (1) social stratification, (2) supervision link prediction, and (3) matching based on the prediction results
  - CREATE-SL: consists of 2 steps (1) social stratification, (2) supervision link prediction
  - CREATE-SM: consists of 2 steps (1) social stratification, and (2) matching based on social links
  - CREATE-S: the social stratification step only
  - Traditional Unsupervised Link Prediction Methods: CN (Common Neighbor), JC (Jaccard's Coefficient), and AA (Adamic Adar)

	CREATE	CREATE-SL	CREATE-SM	CREATE-S	CN, JC, AA
social stratification	√	√	√	√	
link prediction	√	√			√
matching	√		√		

- Evaluation Metrics
  - AUC, Precision@100
  - Precision, Recall and F1 (for CREATE and CREATE-SM)

Social Stratification Experiment Results

- Chart Depth Regulation Constraint Parameter α Selection



Observations:

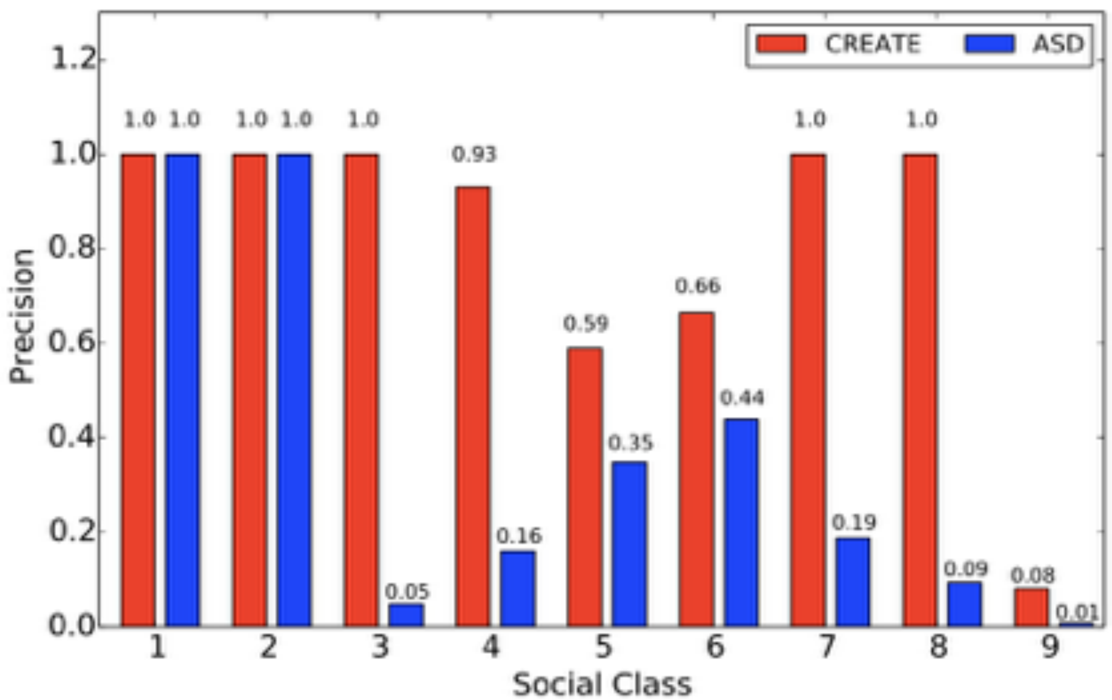
- small α (1-3) has no effects on the social stratification result
- CREATE can achieve the highest accuracy score when α=5.6
- large α (7-9) can lead to very low accuracy scores

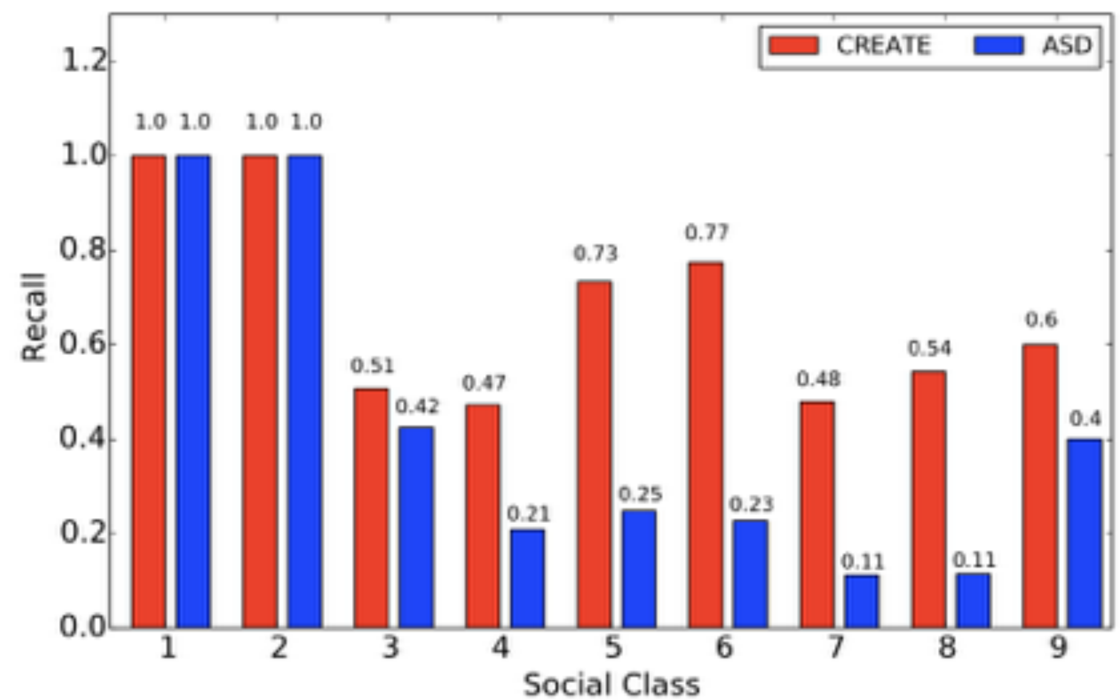
Explanations:

- Matthew Effect based constraint can stretches the organizational chart, the average depth achieved by which is already >3
- α=5.6 is close to the average chart depth of users in ESN
- large α (7-9) stretches the chart to much and place most users at the wrong level

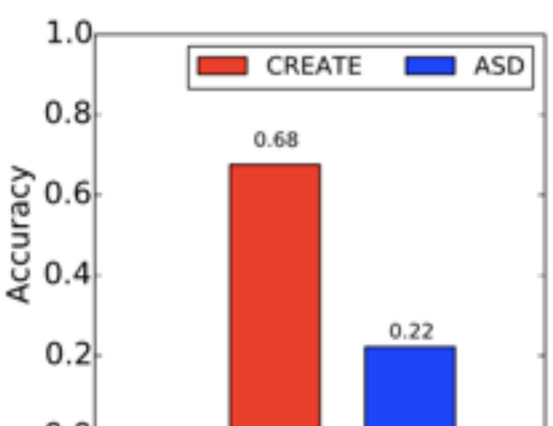
Social Stratification Experiment Results

- Comparison of CREATE and ASD at **EACH** class level

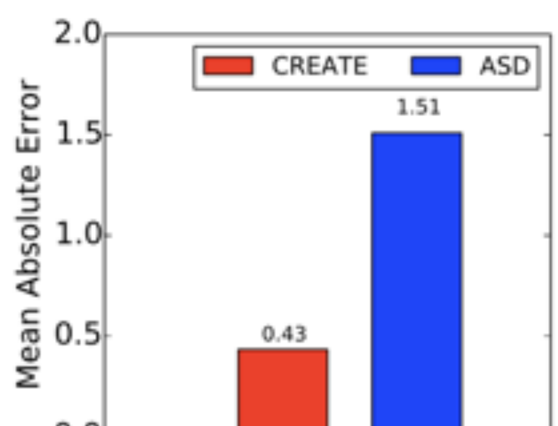




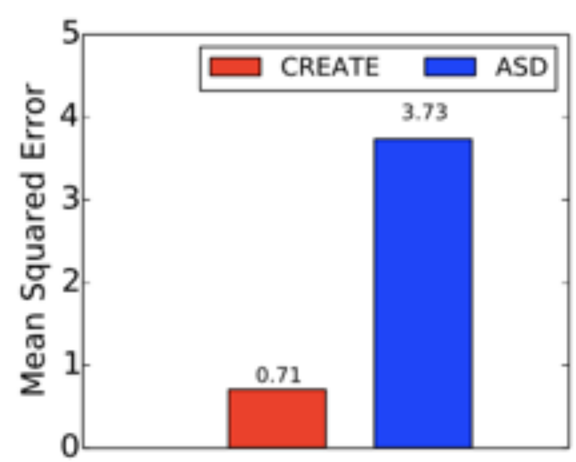
- Comparison of CREATE and ASD at **ALL** class levels



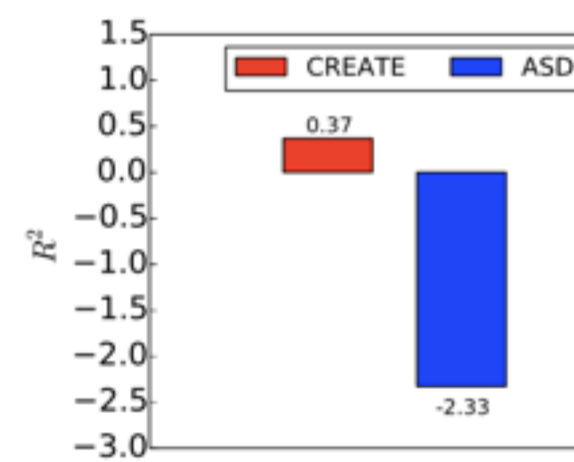
(a) Accuracy



(b) MAE



(c) MSE



(d) R^2

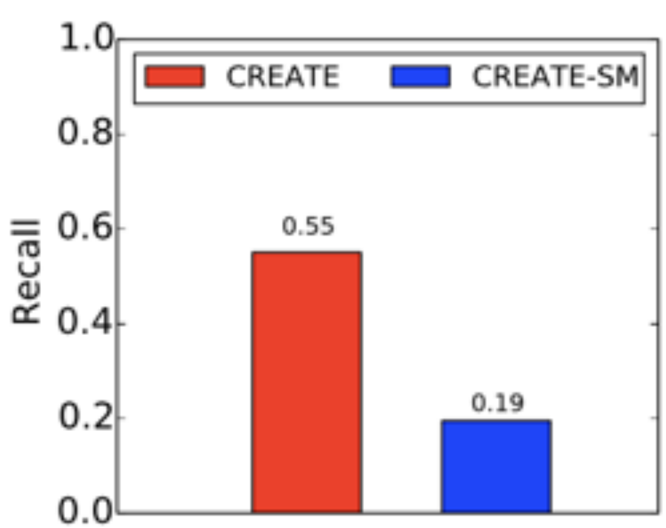
Organizational Chart Inference Experiment Results

Table 1: Performance comparison of different organizational chart inference methods.

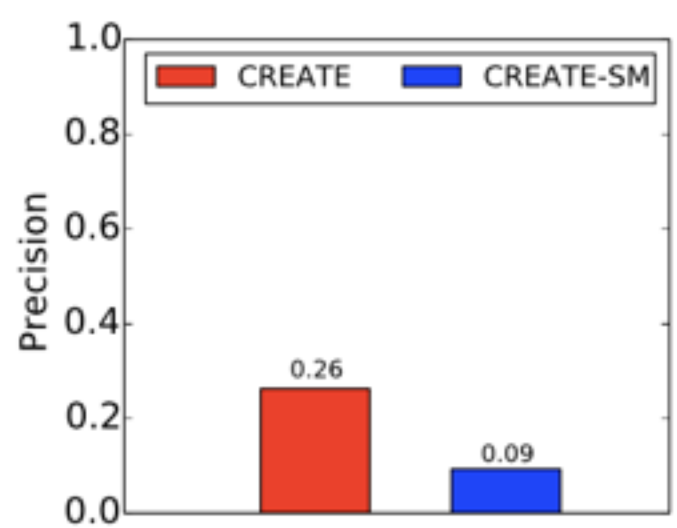
Method	Metrics	
	AUC	Precision@100
CREATE(K = 10)	0.856	0.830
CREATE(K = 15)	**0.869**	**0.870**
CREATE(K = 20)	**0.869**	**0.870**
CREATE-SL	0.719	0.820
CREATE-SM(K = 10)	0.610	0.720
CREATE-SM(K = 15)	0.630	0.790
CREATE-SM(K = 20)	0.630	0.790
CREATE-S	0.627	0.740
S-CN	0.636	0.440
S-JC	0.636	0.260
S-AA	0.528	0.070

Observations:

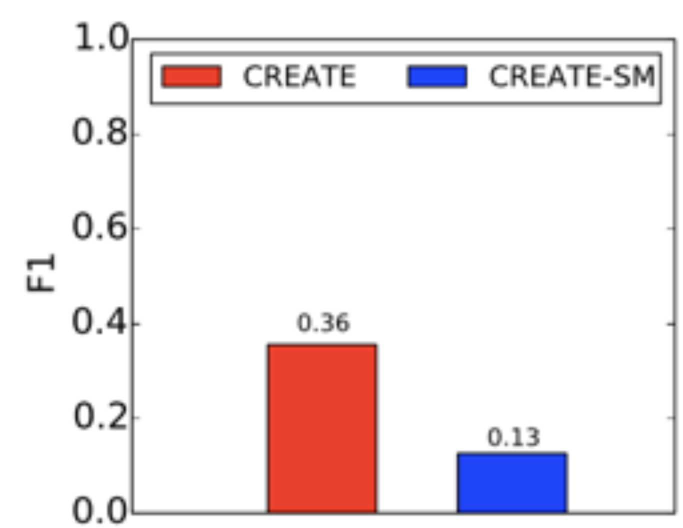
- *CREATE vs CREATE-SL* (*CREATE-SM vs CREATE-S*): Matching step can prune lots of non-existing supervision links
- *CREATE vs CREATE-SM* (*CREATE-SL vs CREATE-S*): Link prediction step can identify more potential supervision links
- *CREATE-SL vs S-CN, S-JC, S-AA*: Link prediction based on meta path aggregation can utilize heterogeneous in ESNs to identify more supervision links among employees.



(a) Recall



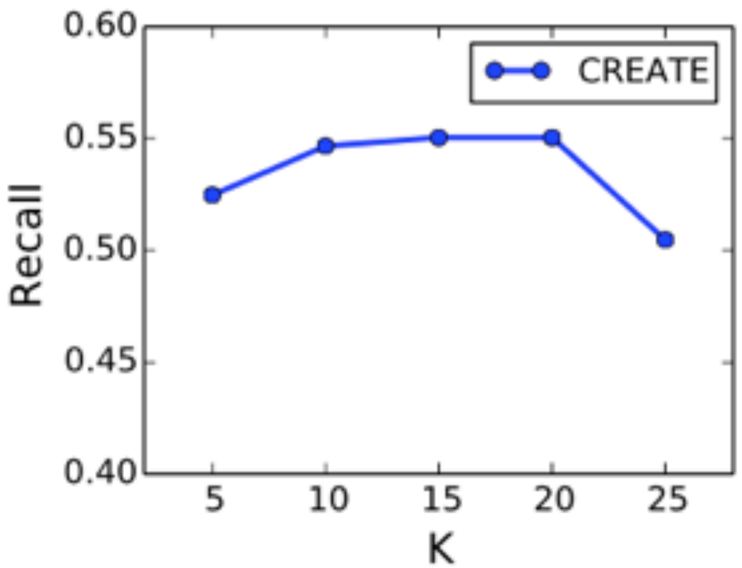
(b) Precision



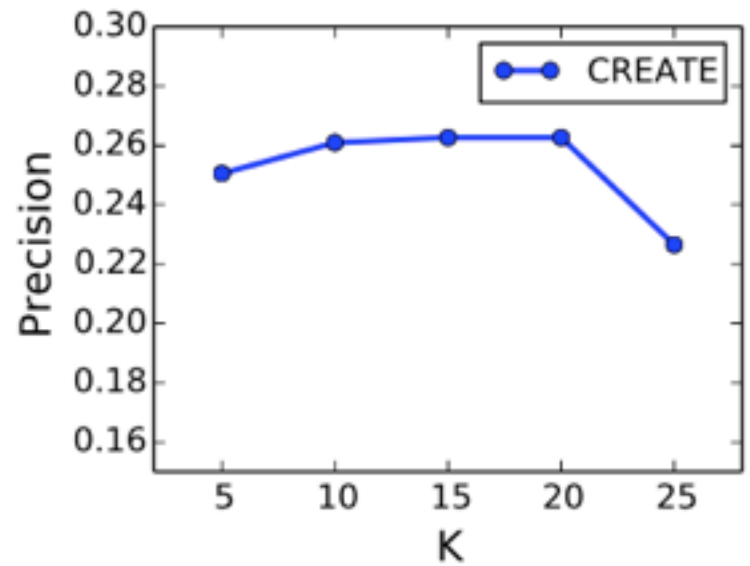
(c) F1

Sensitivity Analysis of Parameters

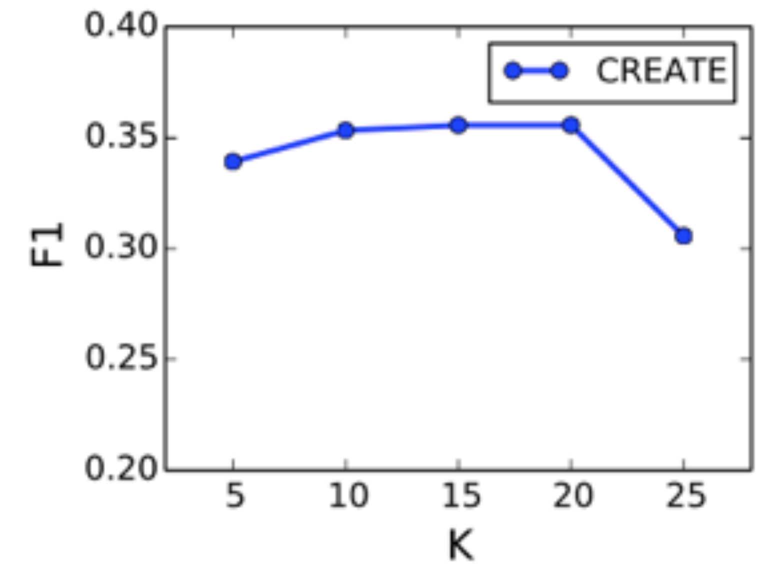
- Consecutive social class K-to-1 matching:



(a) Recall



(b) Precision



(c) F1

Observations:

- small K (5) can lead to bad results
- medium K (10-20) can achieve better performance
- large K (24) can lead to even very bad chart inference results

Explanations:

- Some manager can only supervise small number of employees
- On average, each supervisor can manage around 10-20 subordinates directly
- Few supervisor can have greater than 25 direct subordinates at the same time

Summary

- In this paper, we study the “Organizational Chart Inference” problem
  - Target: recover management hierarchical structures in companies based on information in online enterprise social networks
- Proposed a 3-phrase framework CREATE
  1. social stratification to minimize the class transcendence penalties
  2. supervision link prediction based on meta path aggregation
  3. social class matching based on minimum cost network flow mode
- Conduct extensive experiments on real-world online ESNs and organizational chart (of Microsoft) to test the effectiveness of the CREATE framework

Organizational Chart Inference

Q&A

Jiawei Zhang, Philip S. Yu, and Yuanhua Lv
jzhan9@uic.edu, psyu@cs.uic.edu, yuanhual@microsoft.com