Broad Learning and Deep Social Network Alignment

Lin Meng[1], Yuxiang Ren[1], **Jiawei Zhang[1]**, Fanghua Ye[3], and Philip S. Yu[2]

[1] IFM Lab, Florida State University, FL, USA
[2] University of Illinois at Chicago, IL, USA
[3] Department of Computer Science, University College London, UK

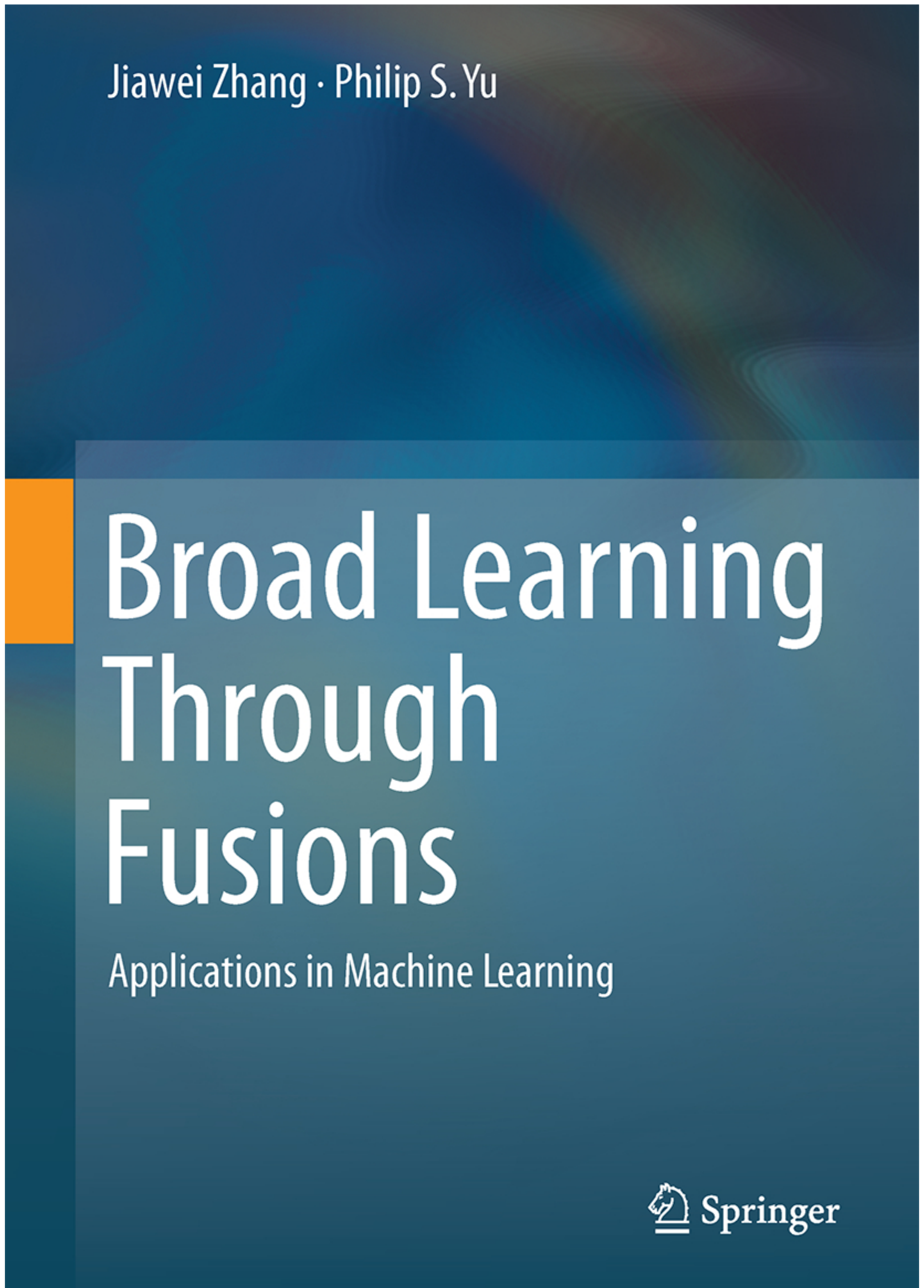
Part 1: Broad Learning

Part 2: This paper

Part 1: Broad Learning

- Broad Learning Definition
- Broad Learning Motivation
- Broad Learning Key Tasks and Challenges
- Broad Learning Examples
- Relation with Other Existing Learning Problems

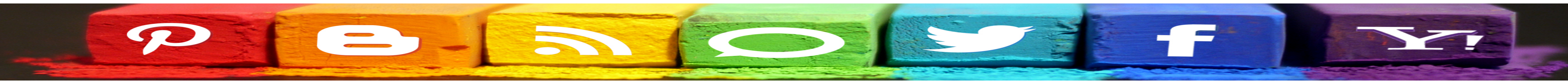
Broad Learning: A Textbook



12 Chapters, 450+ Pages, Available by June 17, 2019
Flyer: https://www.springer.com/us/book/9783030125271

Table of Contents

Broad Learning



Broad Learning Definition

- ***Broad Learning*** is a new type of learning task, which focuses on fusing multiple large-scale information sources of diverse varieties together and carrying out synergistic learning tasks across these fused sources in one unified analytic.

Motivations

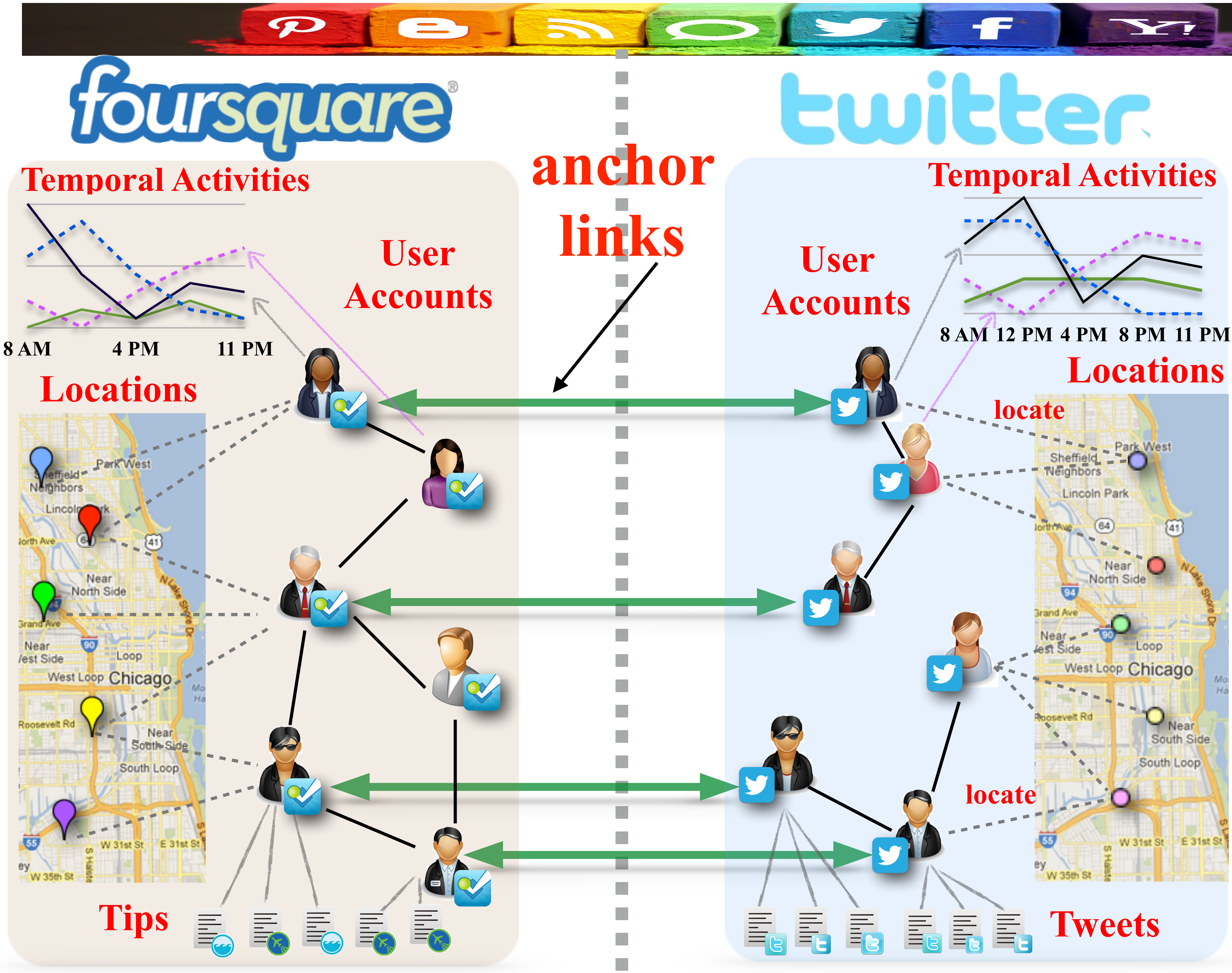
- In the real world, on the same information entities, a large amount of information can actually be collected from various sources.
- These sources are usually of different varieties. Each information source provides a specific signature of the same entity from a unique underlying aspect.
- Effective fusion of these different information sources provides an opportunity for researchers and practitioners to understand the entities more comprehensively, which renders “Broad Learning” an extremely important learning task.

Motivation: Why Fusing Multiple Information Sources

Broad Learning Key Tasks and Challenges

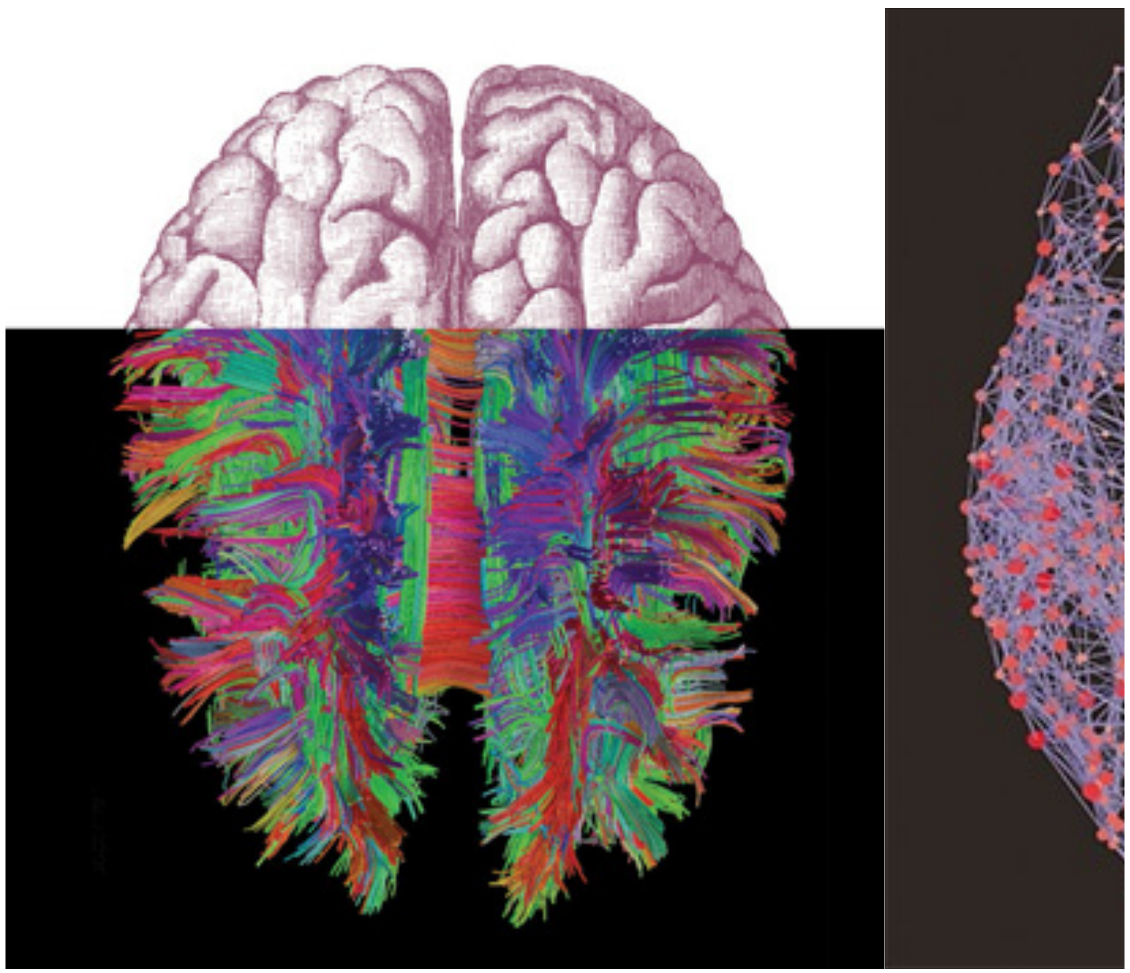
- **Broad Learning Key Tasks**
  - ***Information Fusion:*** effective fusion of information across multiple data sources
  - ***Knowledge Learning:*** effective learning of the multiple fused data sources
- **Broad Learning Key Challenges**
  - How to Fuse?
    - Data fusion strategy will vary for different types of data input
    - e.g., Networks, Images, Text, Speech, etc.
  - How to Learning?
    - Not all data will be useful for certain learning tasks
    - Pick the useful information in the learning process

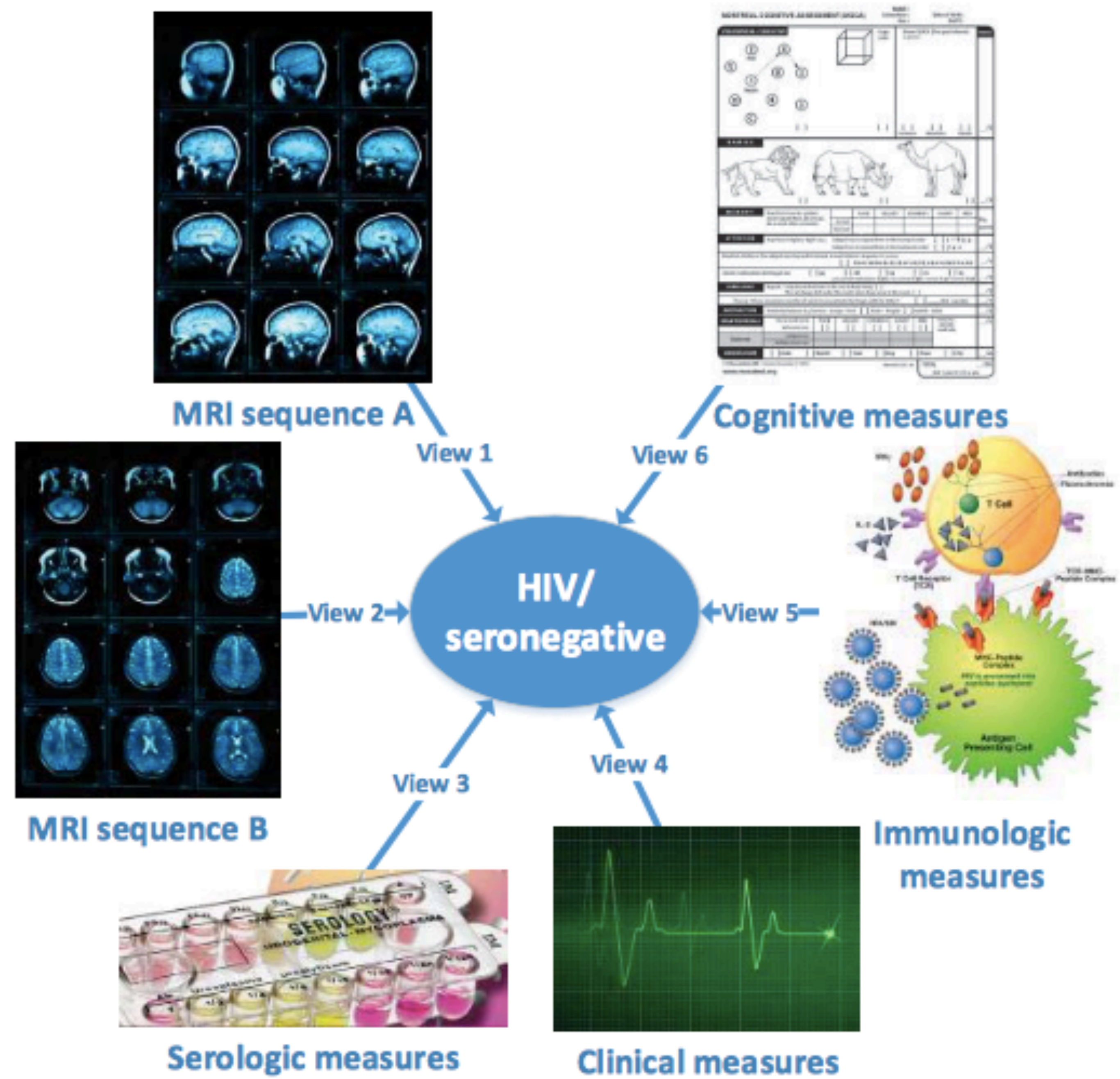
Broad Learning Application in Social Networks Studies



Broad Learning Application in Bio-Medical Studies

Patient Disease Prediction



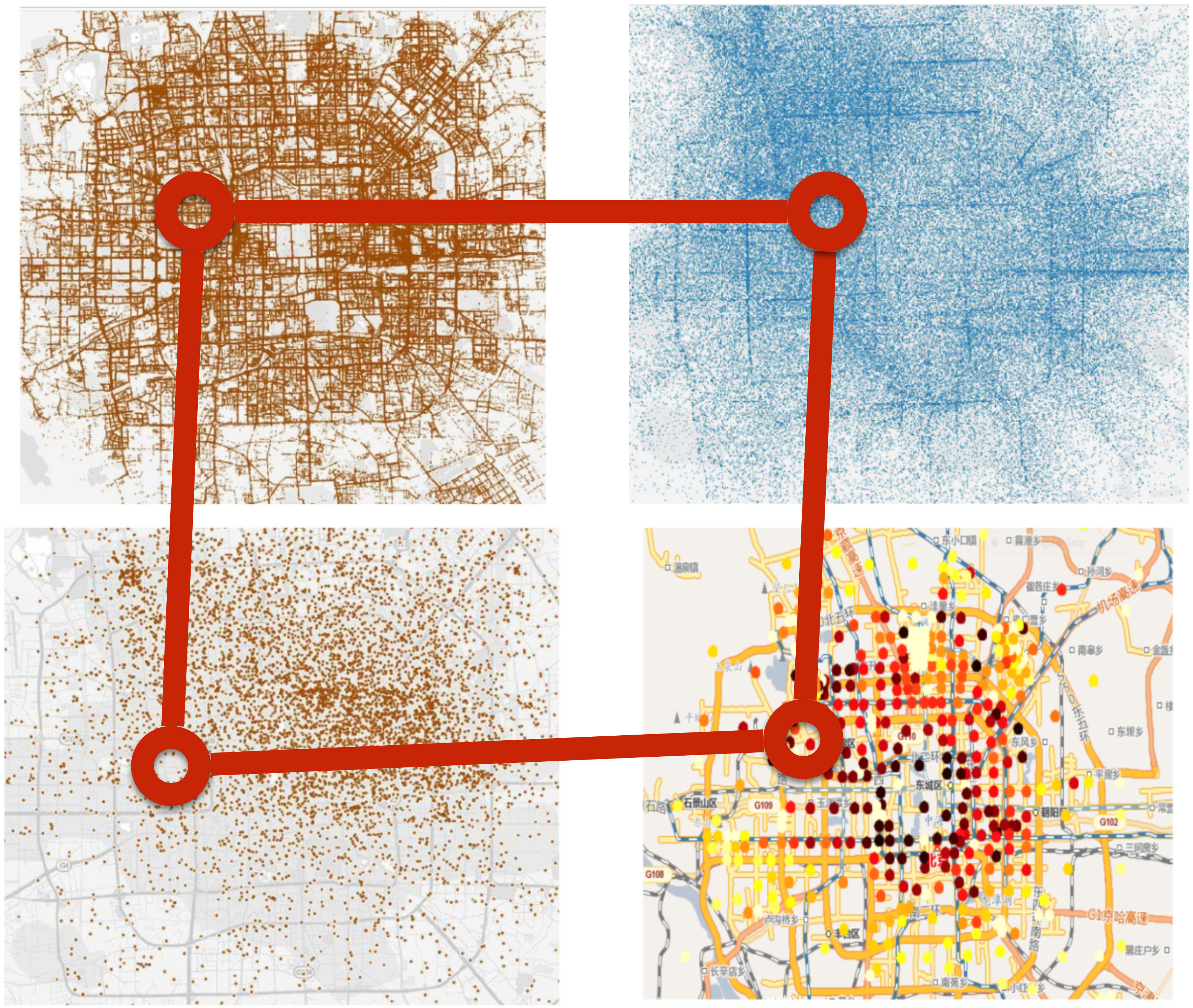


Broad Learning Application in Smart City Research

Smart City

Taxi drop-off points

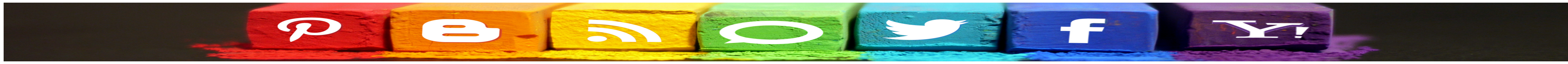
Bus drop-off points



Check-ins

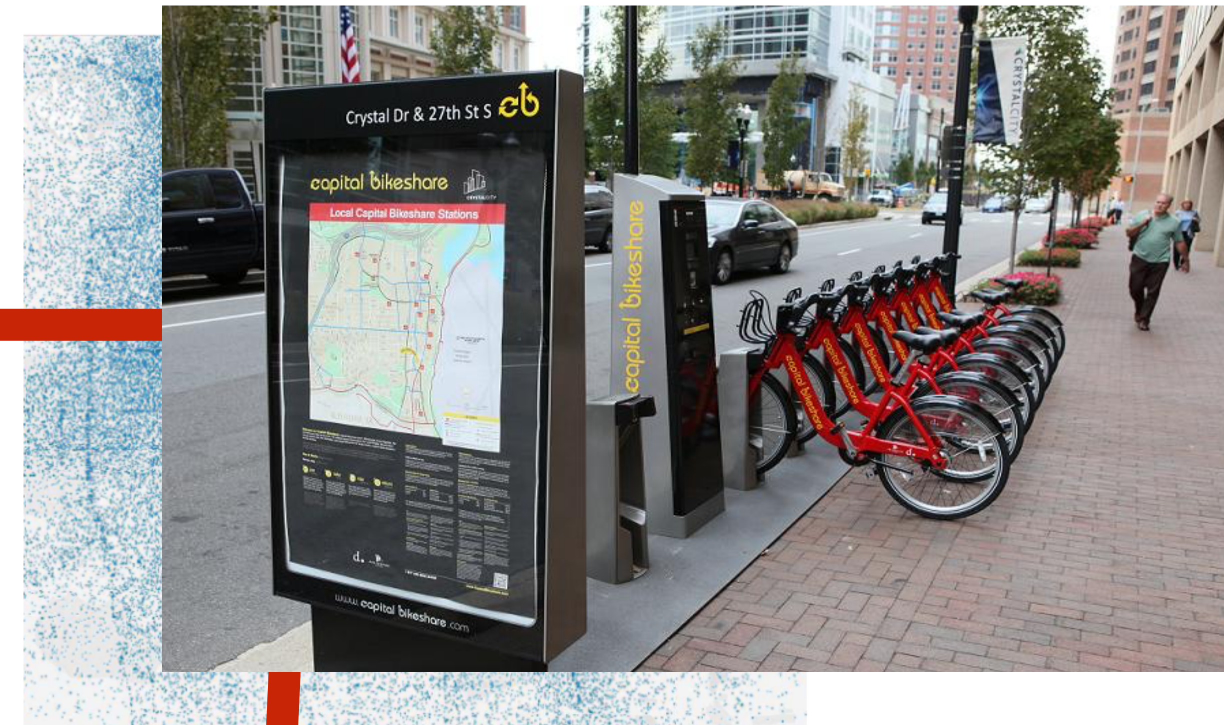
Housing returns

Broad Learning Application in Smart City Research



Gas Station Site Selection

Bike Station Expansion



ff

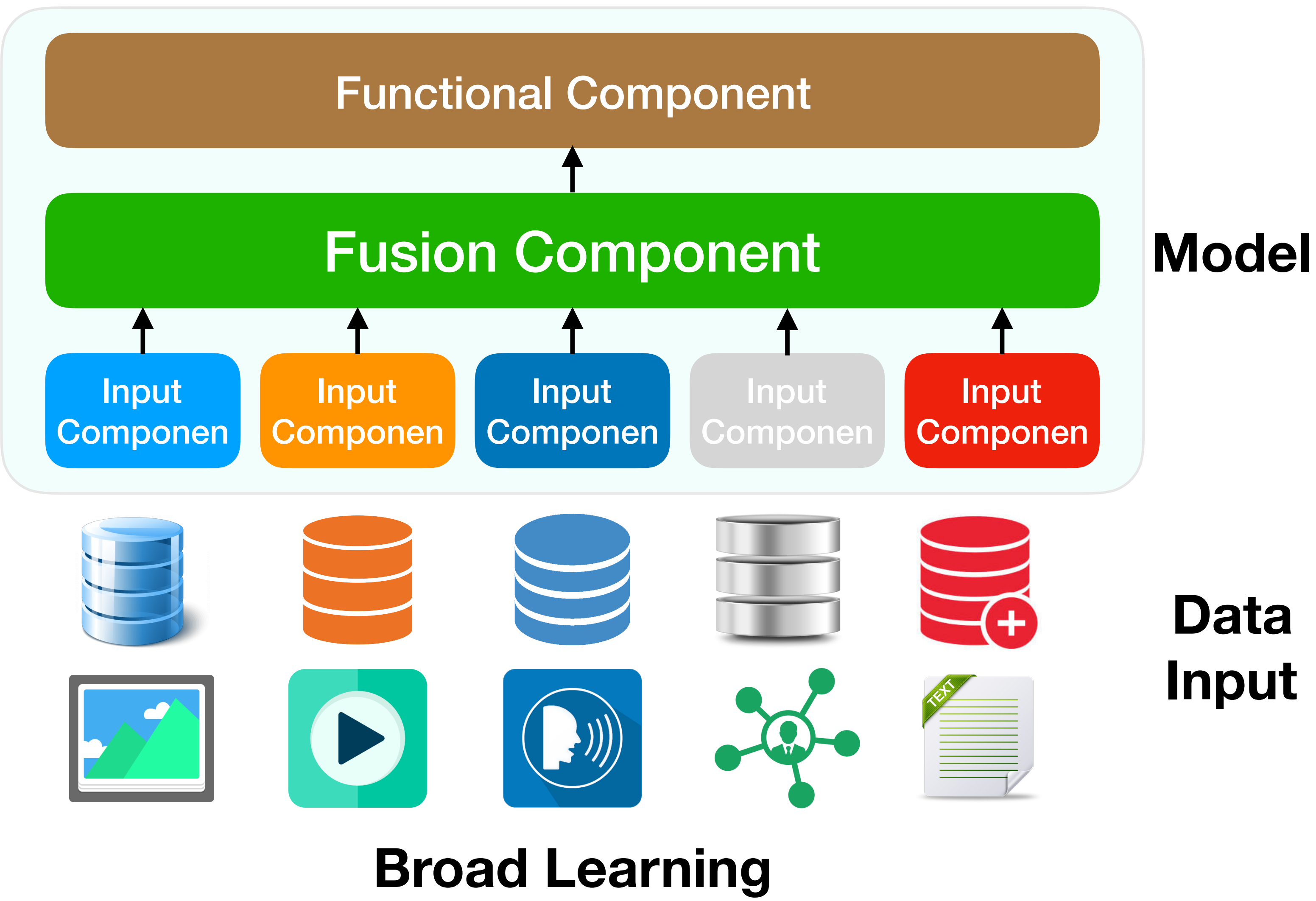
Check-ins

Real Estate Ranking and Planning

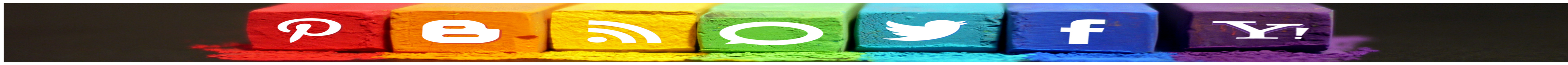
Housing returns

General Broad Learning Architecture

- ***Input Component:*** handles diverse inputs for representation learning
- ***Fusion Component:*** integrate the diverse input via certain strategies
- ***Functional Component:*** performs specific functions, e.g., classification, regression, and clustering, etc.

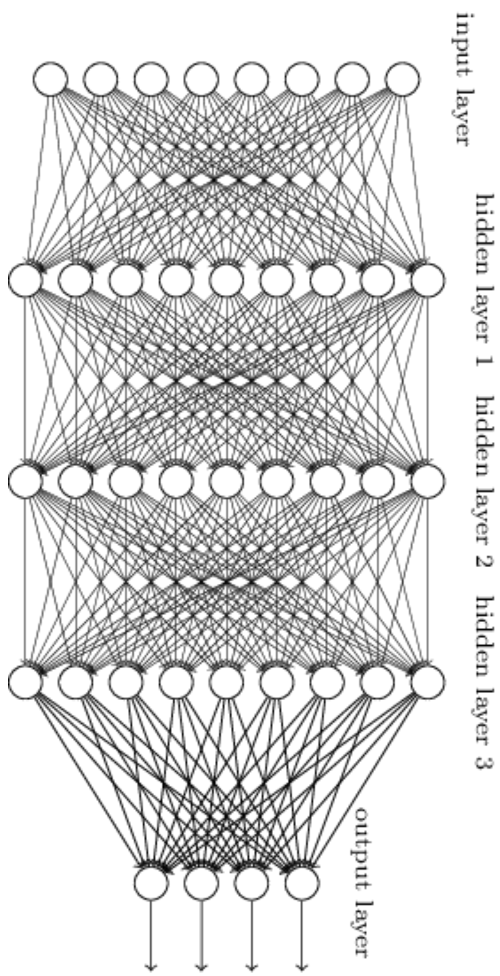


Comparison with Existing Learning Tasks

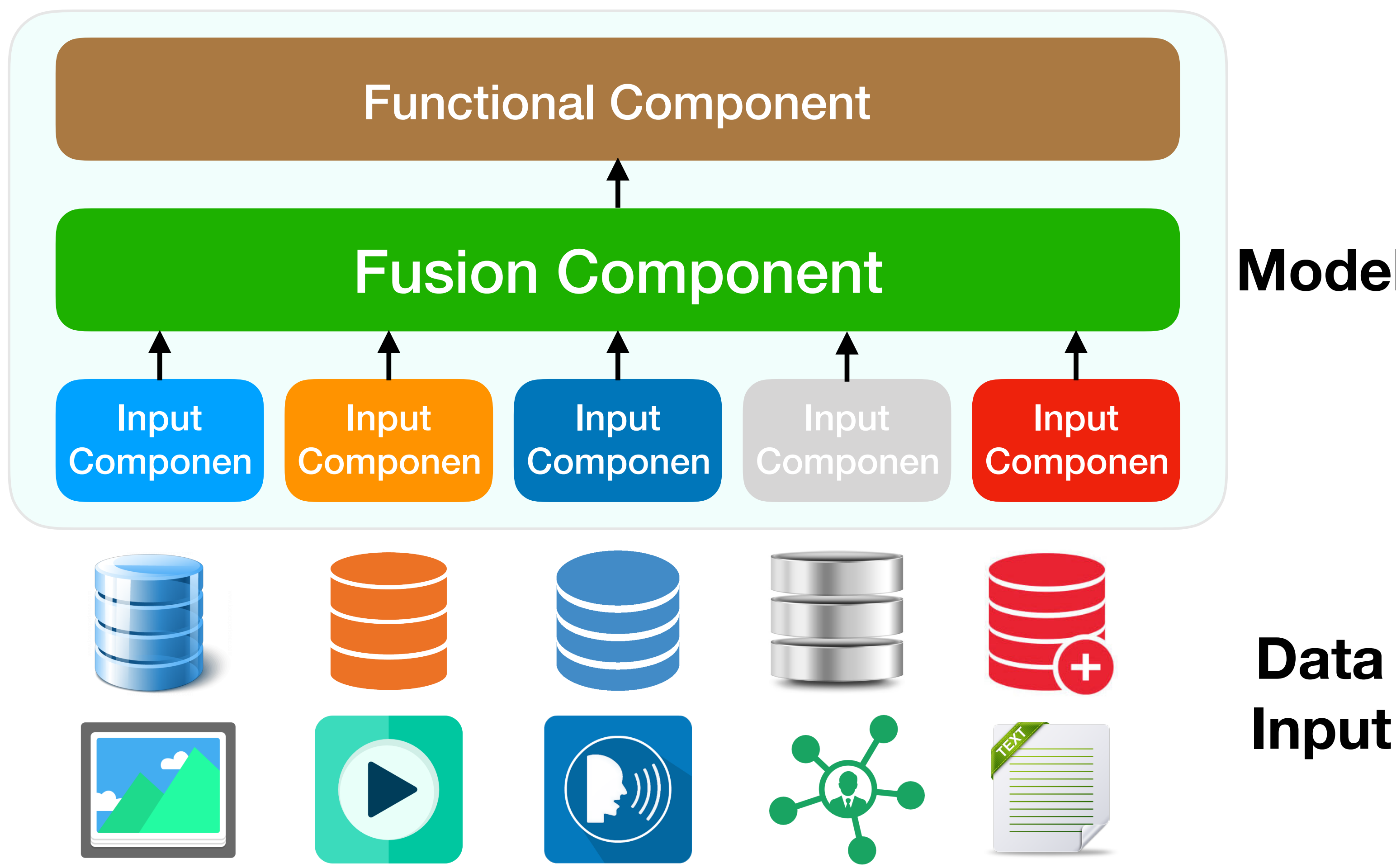


Broad Learning vs Deep Learning

- ***Broad Learning*** is Broad in terms of both data input and the model components.
- ***Deep Learning*** is Deep mainly in terms of model layers.



Deep Learning

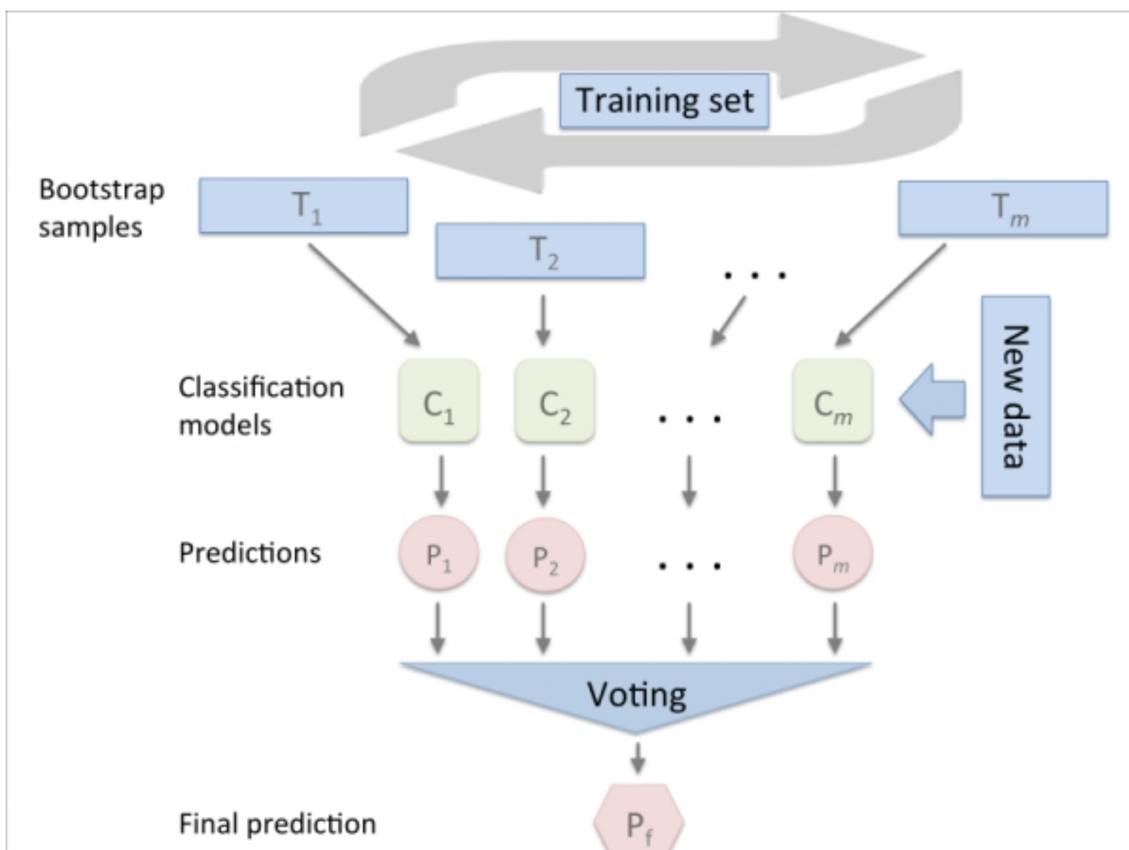


Broad Learning

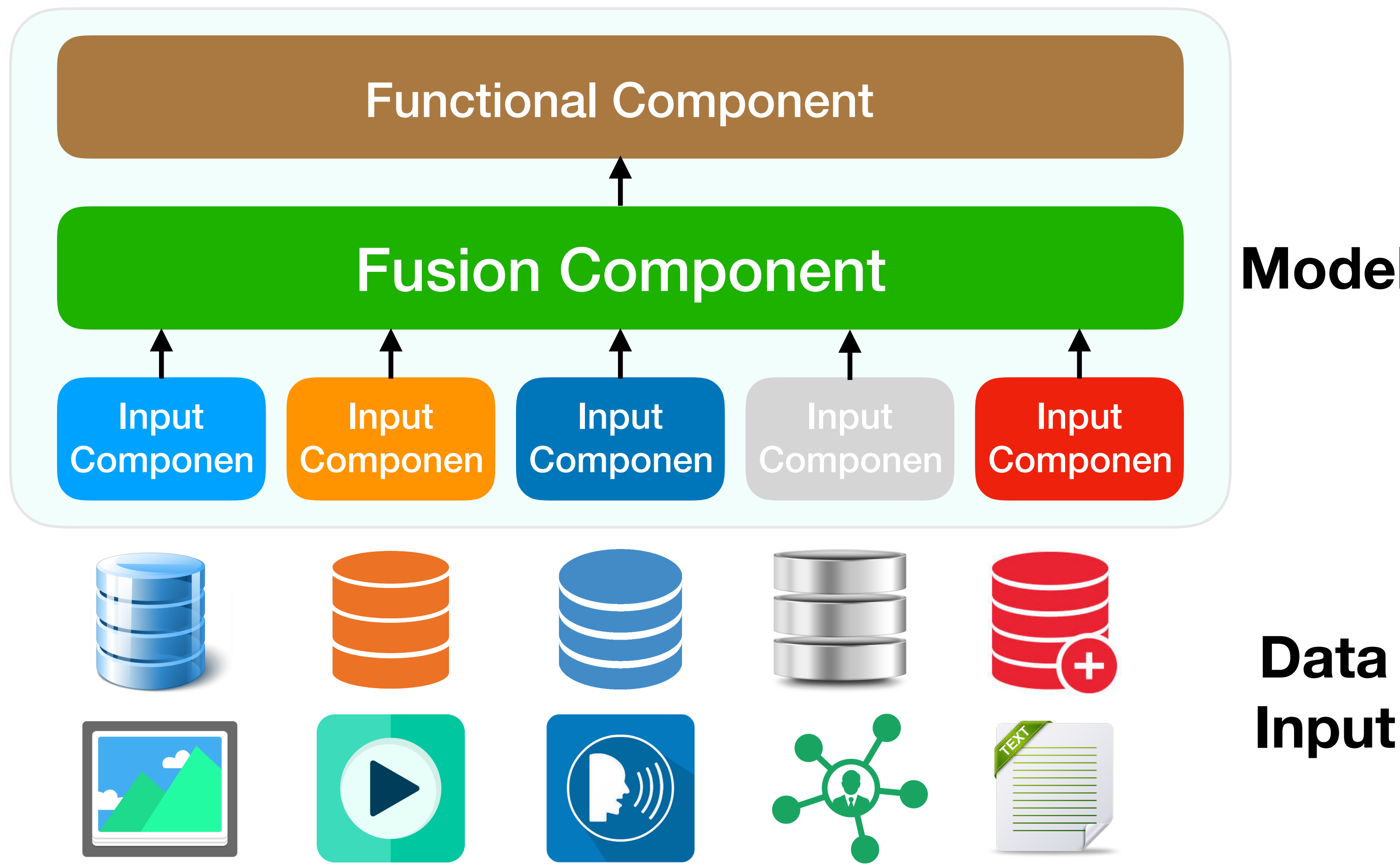
Comparison with Existing Learning Tasks

Broad Learning vs Ensemble Learning

- ***Broad Learning*** combines knowledge from multiple data sources
- ***Ensemble Learning*** combines unit models trained based on the same data source



Ensemble Learning



Broad Learning

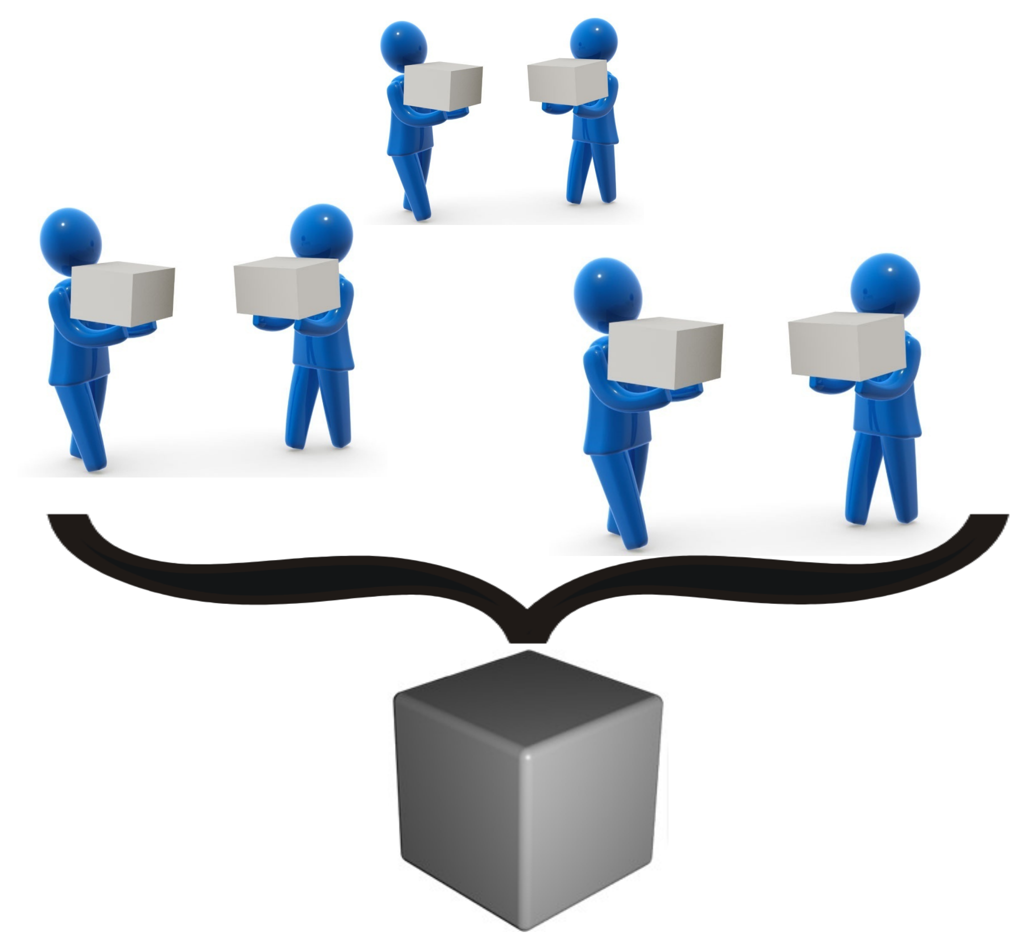
Comparison with Existing Learning Tasks

Broad Learning vs Transfer Learning

- ***Broad Learning*** aims at integrating of information across multiple sources.
- ***Transfer Learning*** focuses on immigration of information between sources.

Transfer Learning is a Special Case of Broad Learning

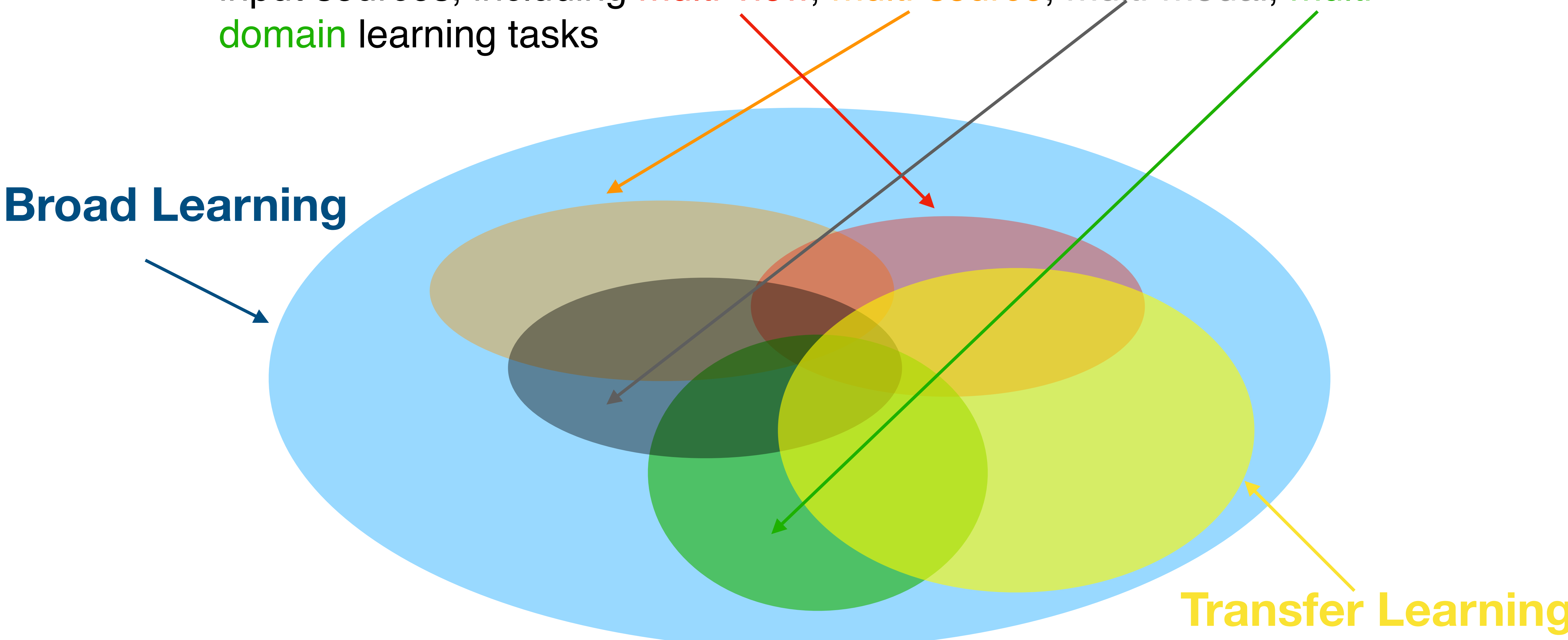
Transfer Learning



Broad Learning

Comparison with Existing Learning Tasks

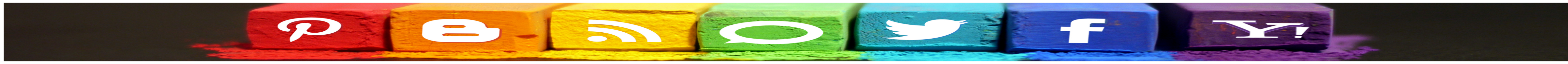
- **Broad Learning vs Multi-View, Multi-Source, Multi-Modal, Multi-Domain Learning**
  - ***Broad Learning*** covers all the existing learning tasks with multiple input sources, including multi-view, multi-source, multi-modal, multi-domain learning tasks

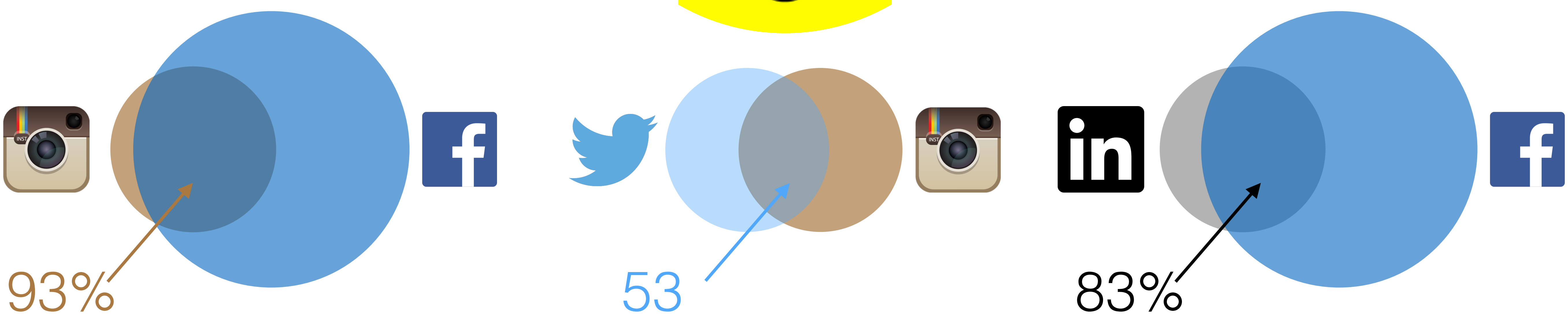


Part 2: This Paper

- Background Knowledge
- Challenges & Solutions
- Framework Architecture
- Experimental Results
- Summary

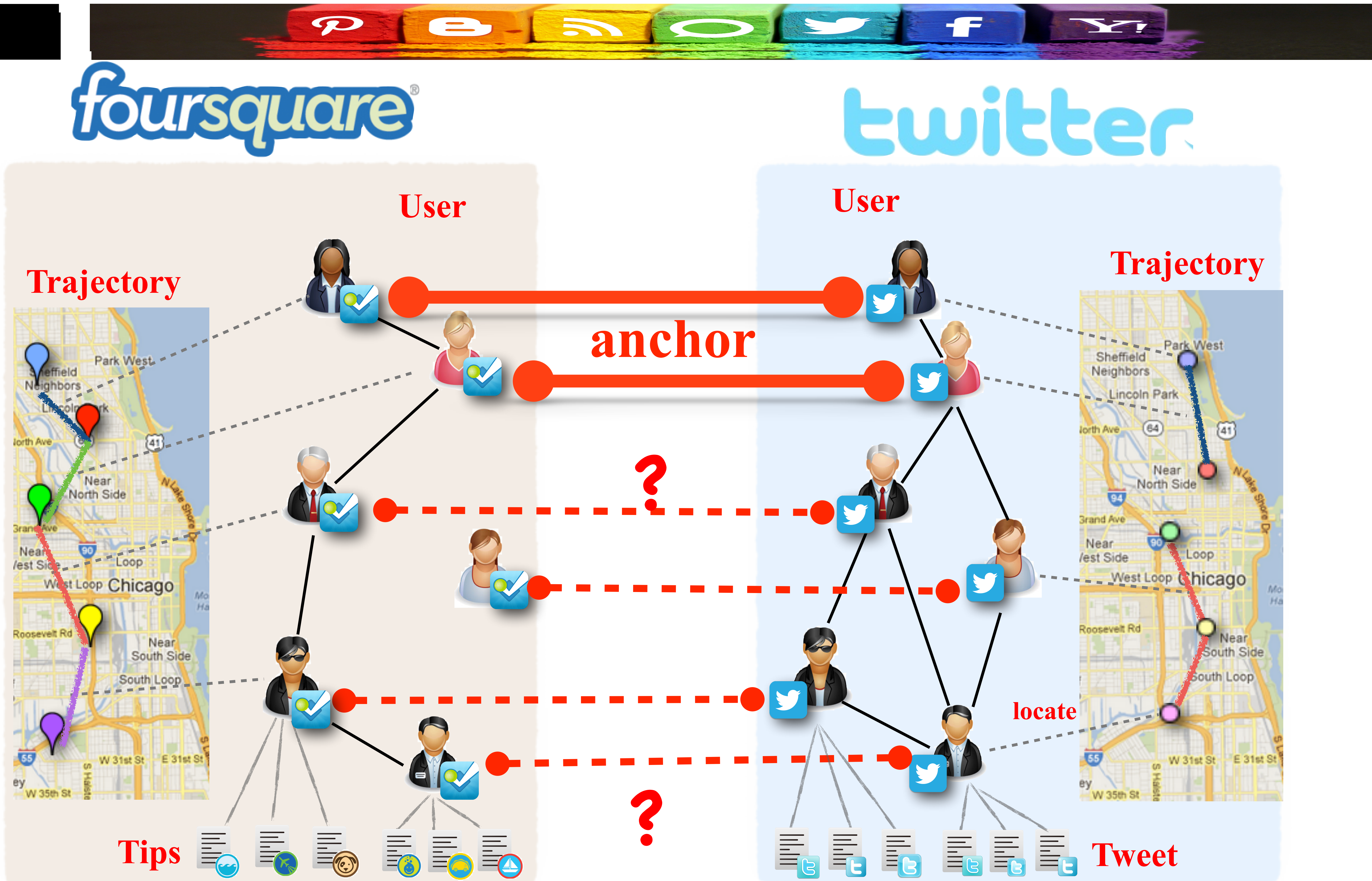
An Observation: People are using multiple online social networks simultaneously



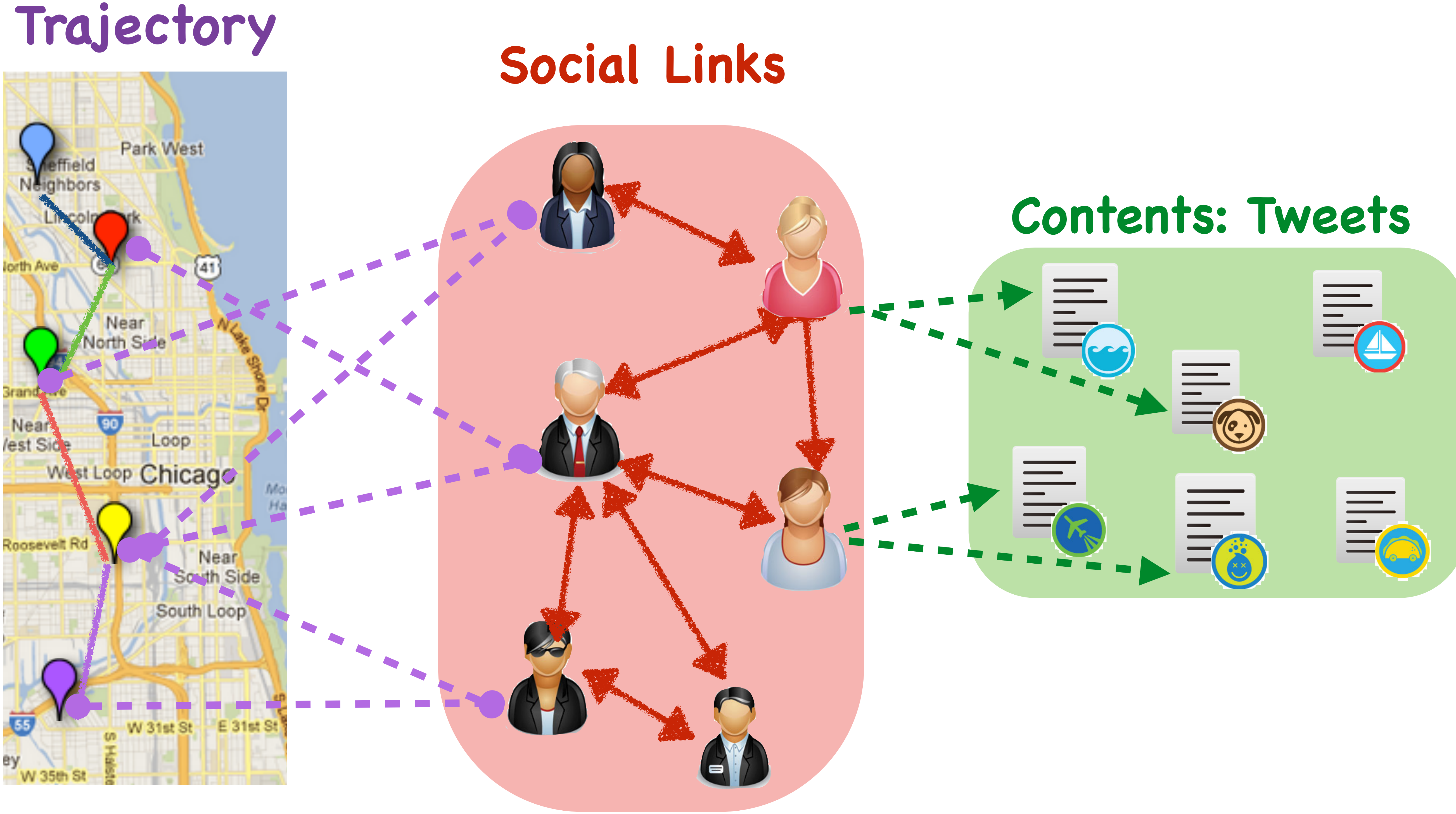


[1] Duggan et al. Social media update, 2013.

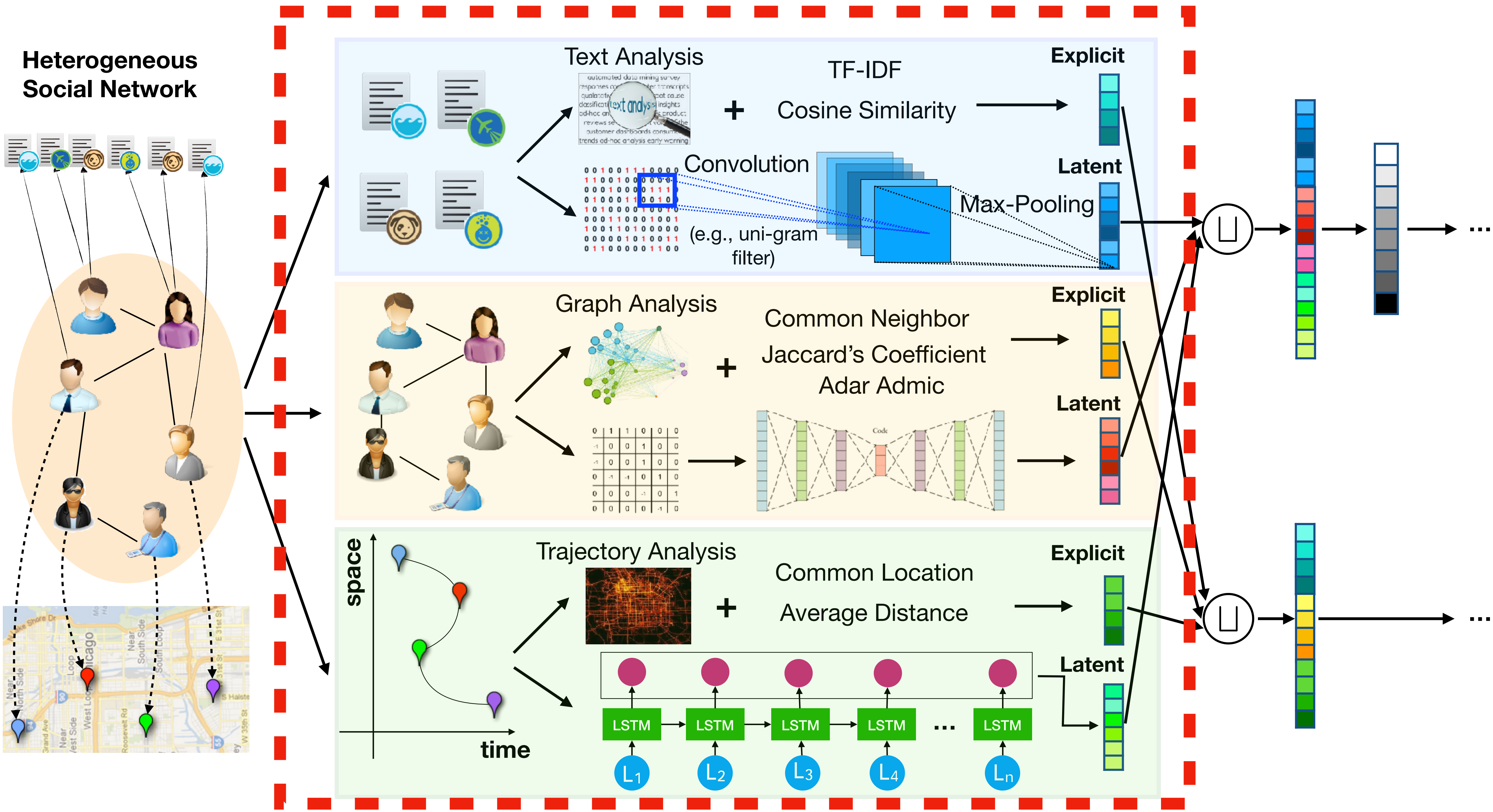
Problem Studied: Heterogeneous Social Network Alignment



Challenge I: Network Heterogeneity

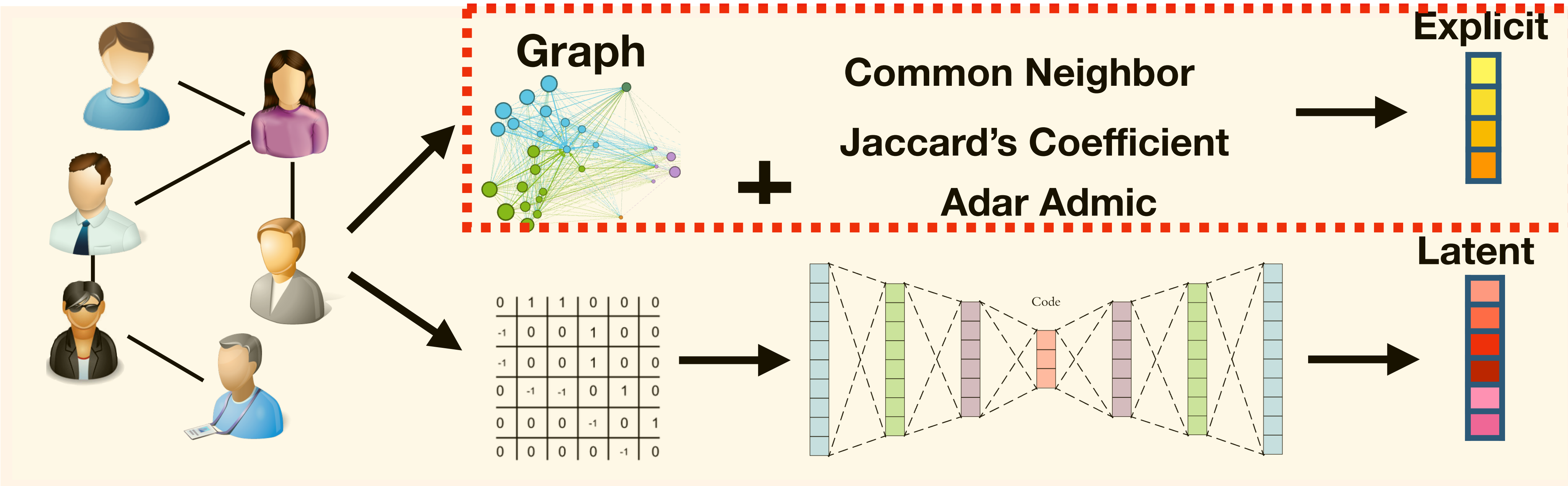


Proposed Method: Heterogeneous Social Networks => Three Representation Modules for Three Types of Data Entities



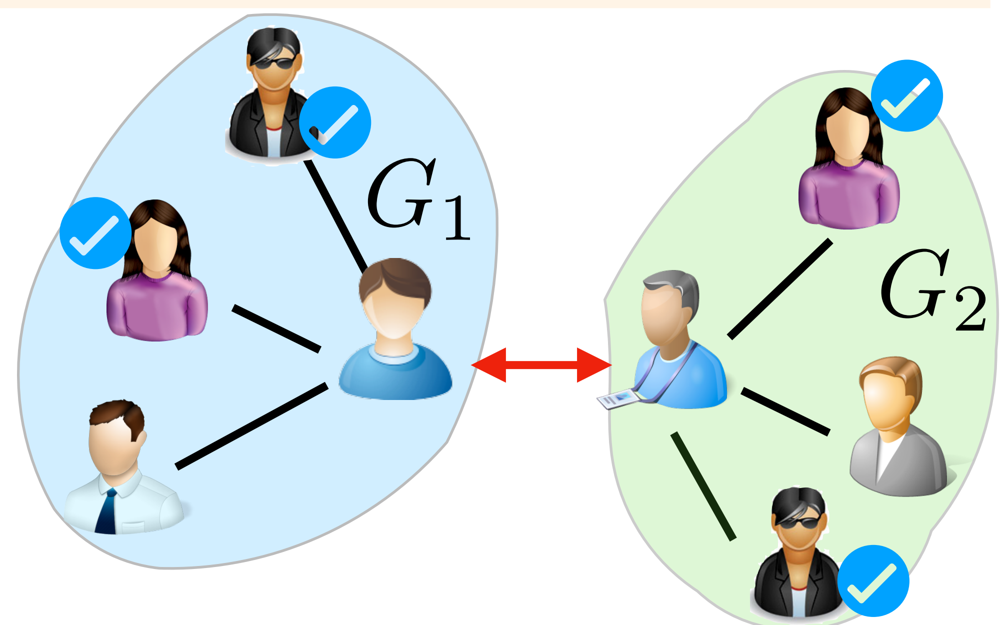
Proposed Method: (1) Social Connection Representation Learning

a. Explicit feature extraction



- Extended common neighbors
- Extended Jaccard's coefficient
- Extended Adamic/Adar

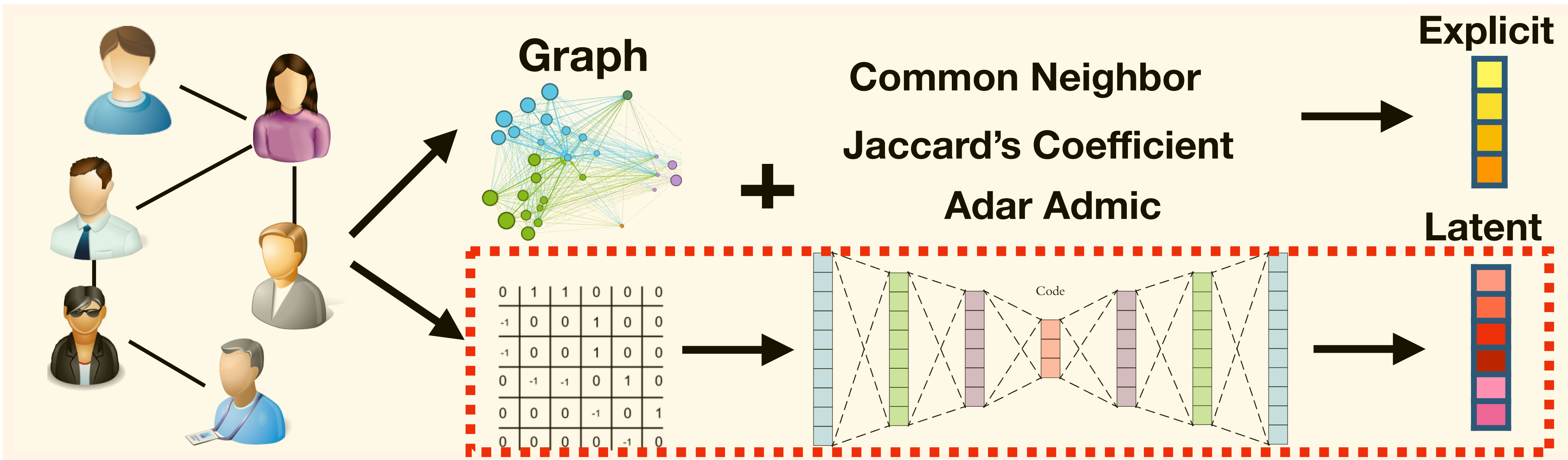
Explicit features are based on anchor links for pairwise users.



Extended common neighbors = 2

Proposed Method: (1) Social Connection Representation Learning

b. Latent feature extraction with Autoencoder model



- **First order proximity**

$$\mathcal{L}_1 = \sum_{u_i, u_j \in \mathcal{U}} S_{i,j} \cdot \| \mathbf{z}_i - \mathbf{z}_j \|_2^2 = Tr(\mathbf{Z}^\top \mathbf{L} \mathbf{Z})$$

Objective function:

$$\mathcal{L}_{SNE} = \beta \cdot \mathcal{L}_1 + \eta \cdot \mathcal{L}_2$$

scalar weights

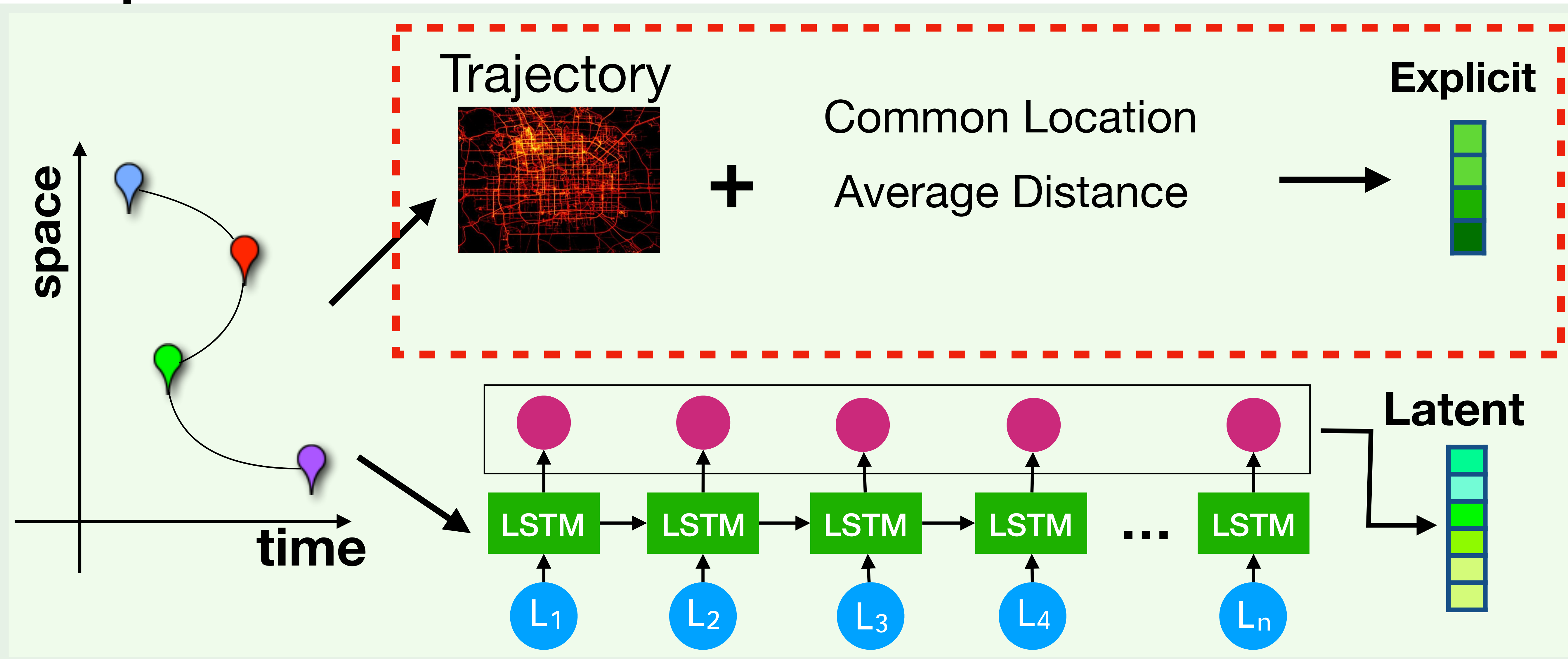
- $\mathbf{L}$ is the laplacian matrix of S and $S_{i,j} = \begin{cases} 1, & \text{if } u_i \text{ and } u_j \text{ are connected;} \\ -1, & \text{if } u_i \text{ and } u_j \text{ are not connected.} \end{cases}$

- **Second order proximity**

$$\mathcal{L}_2 = \sum_{u_i \in \mathcal{U}} \| (\mathbf{x}_i - \hat{\mathbf{x}}_i) \otimes \mathbf{b}_i \|_2^2 \quad \text{where } b_i(j) = \begin{cases} \alpha, & \text{if } x_i(j) = 1; \\ 1, & \text{if } x_i(j) = 0. \end{cases}$$

Proposed Method: (2) Trajectory Representation Learning

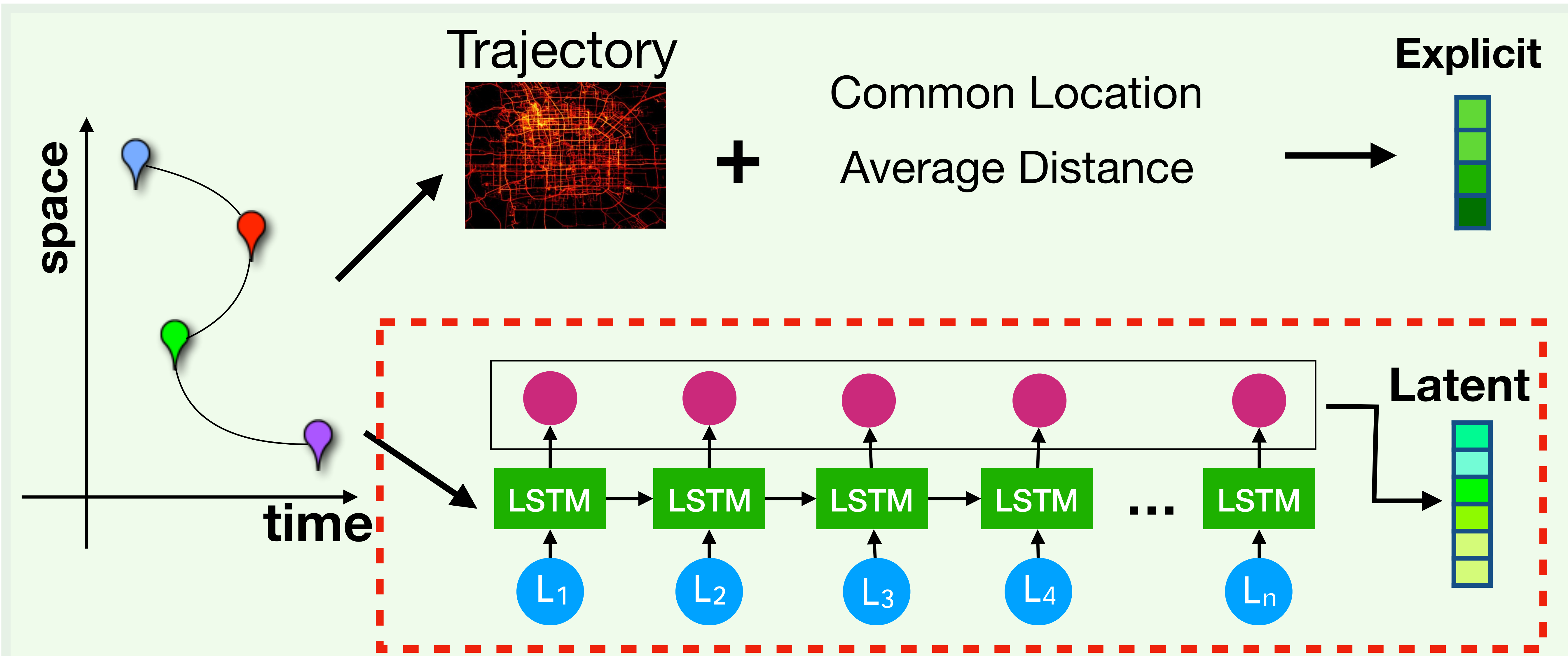
a. Explicit feature extraction



- Number of Common Location
- Average Distance

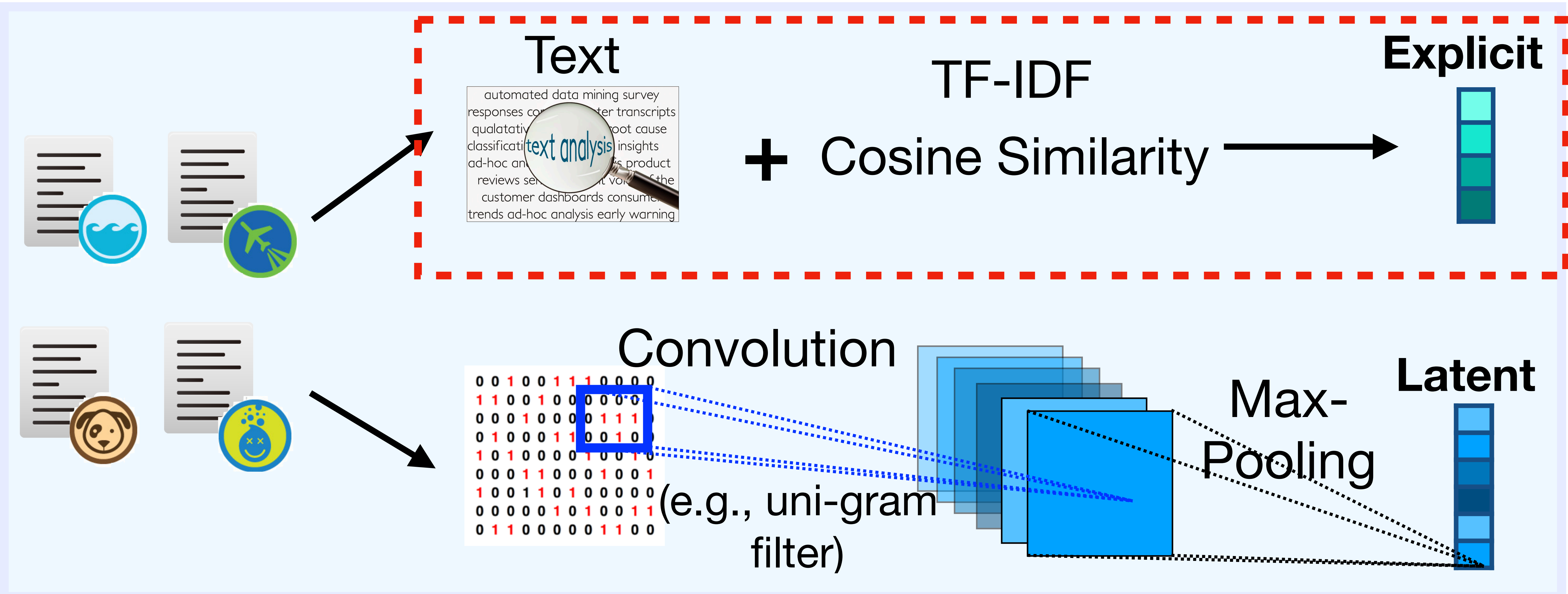
Proposed Method: (2) Trajectory Representation Learning

b. Latent feature extraction with LSTM model



Proposed Method: (3) Textual Word Representation Learning

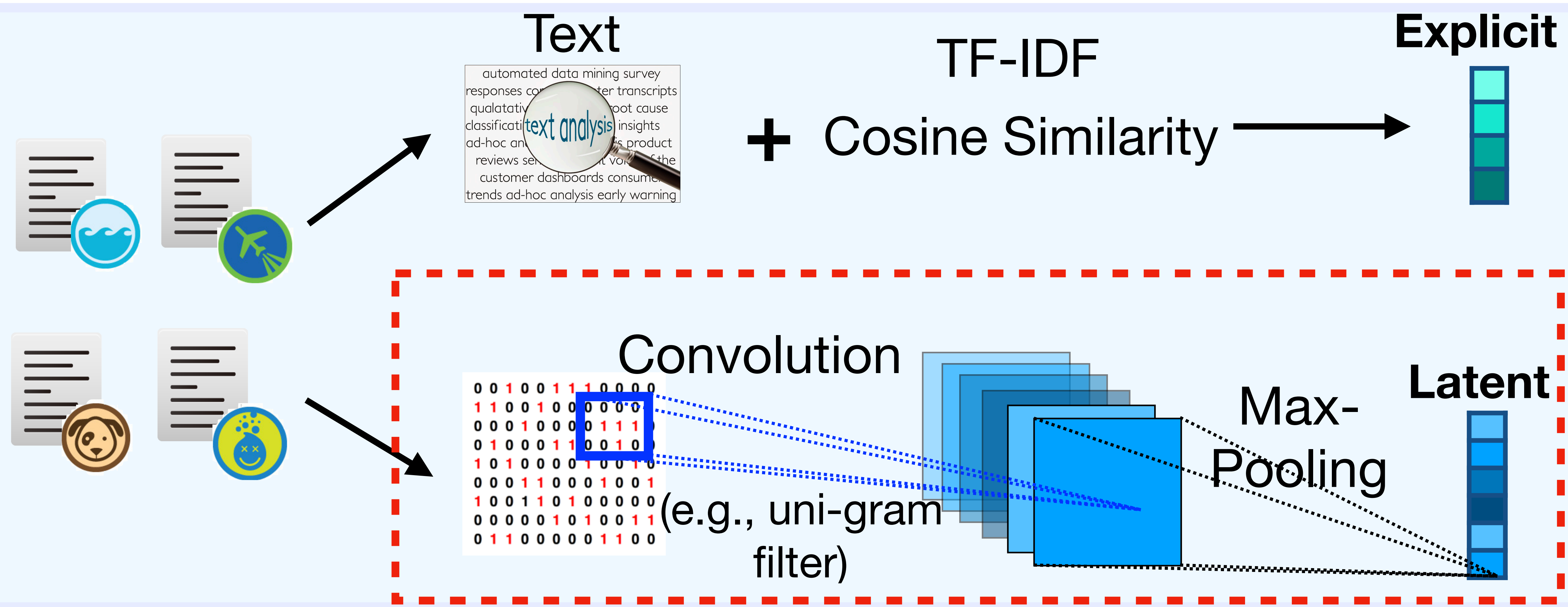
a. Explicit feature extraction



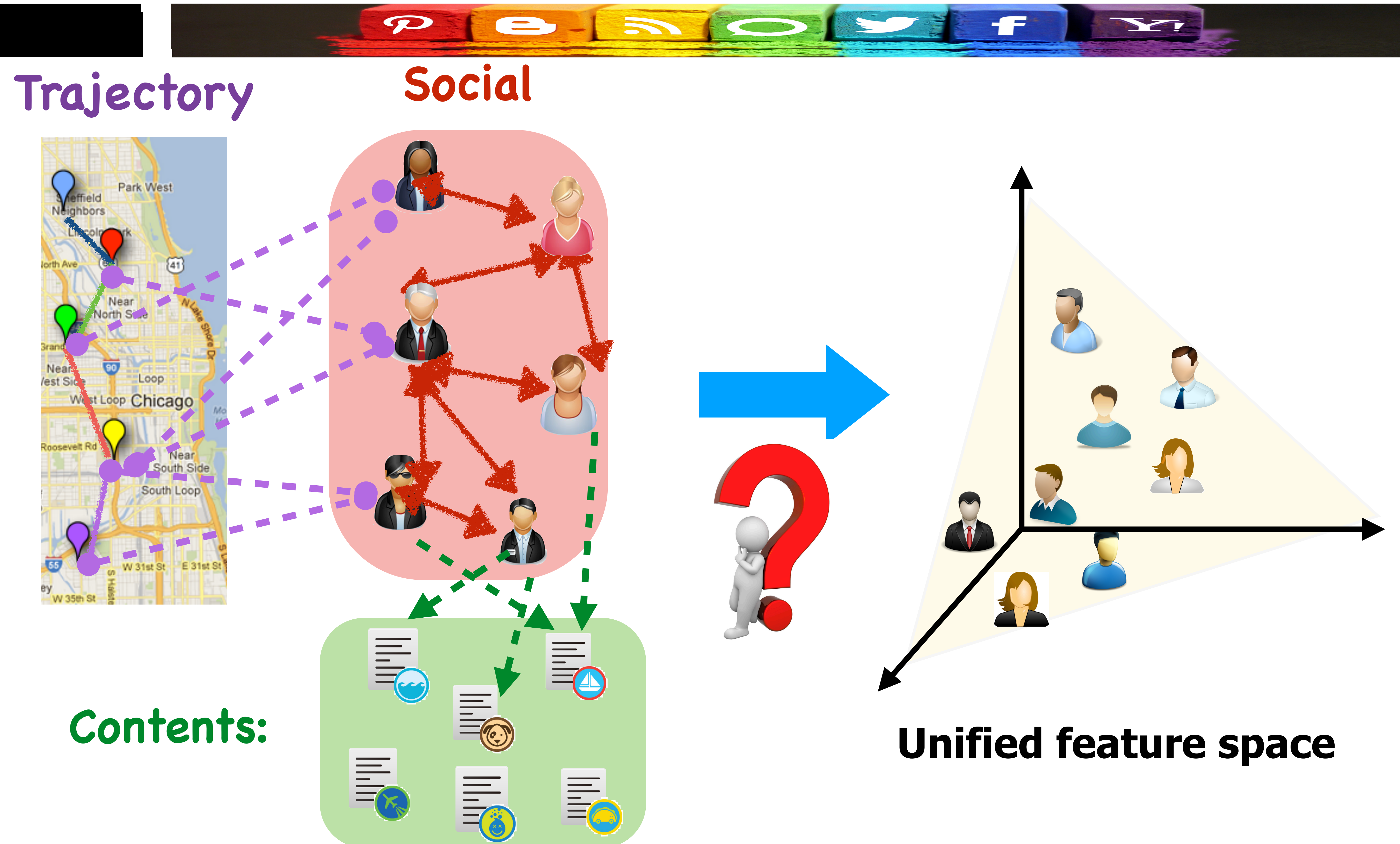
- Extended TF-IDF: two normalized common words

Proposed Method: (3) Textual Word Representation Learning

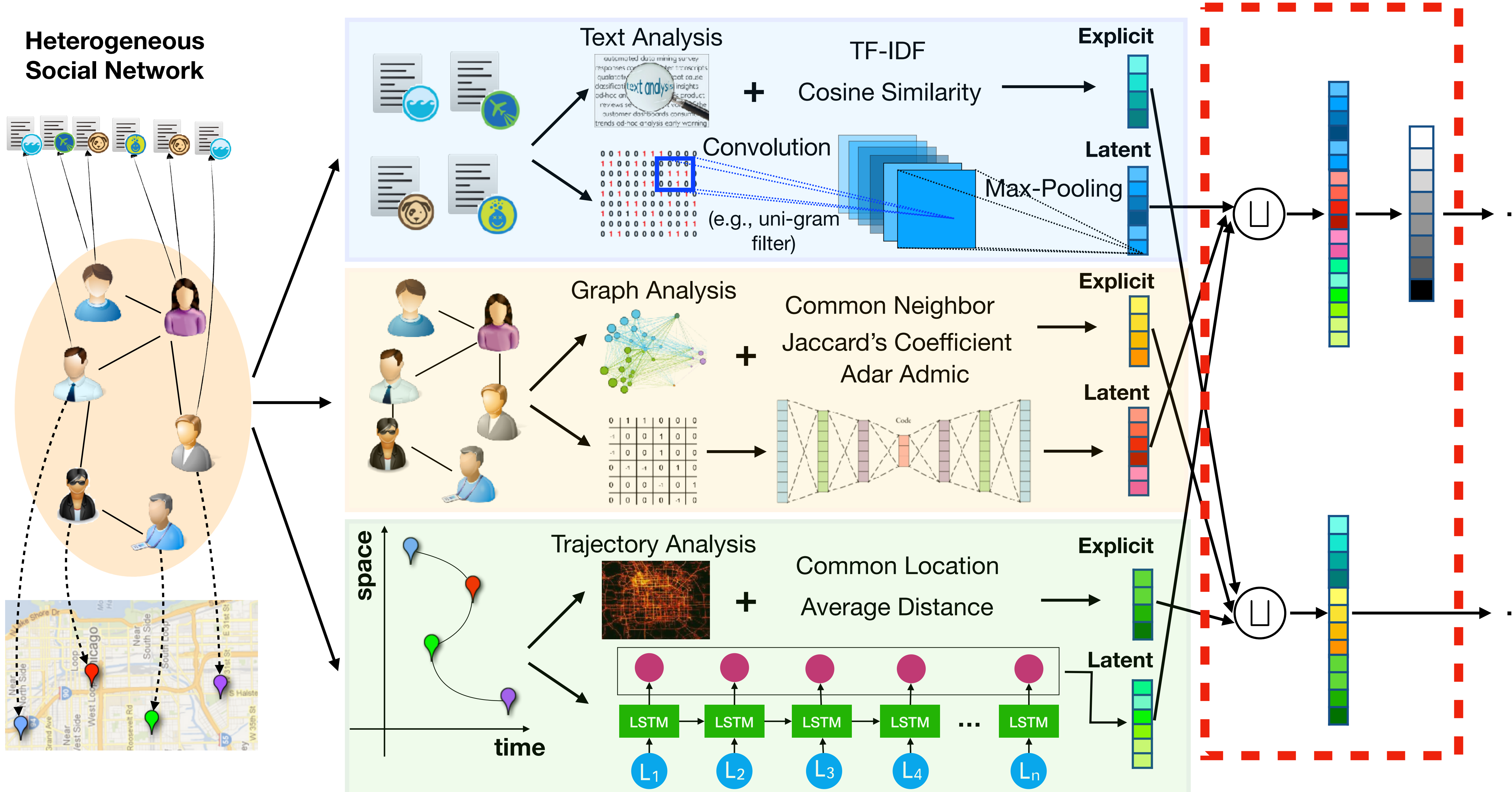
b. Latent feature extraction with CNN model



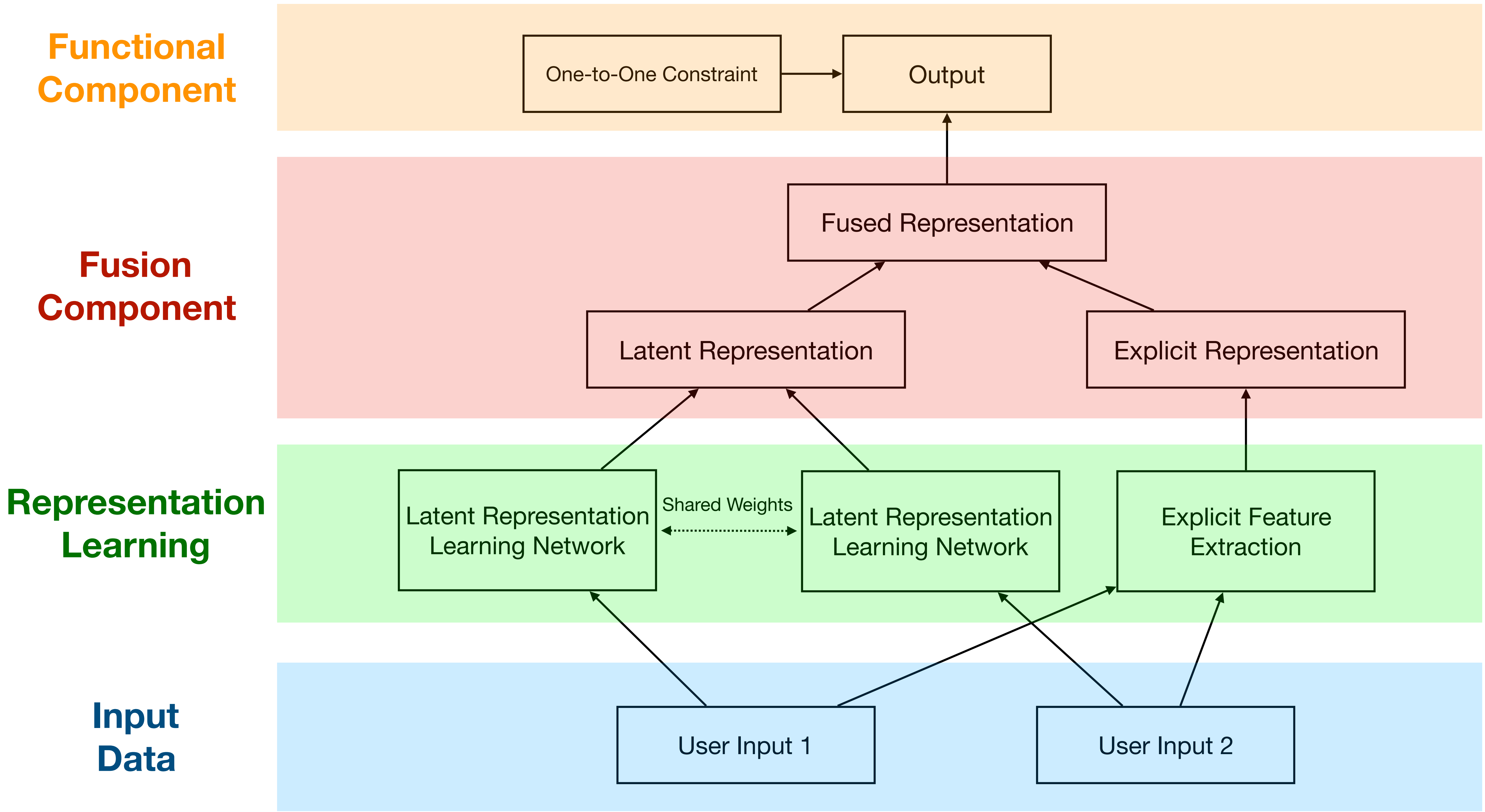
Challenge 2: Information Integration



Proposed Method: Feature Representation Fusion



Proposed Method: DETA Model Framework



Proposed Method: Joint Objective Function

$$\min_{\mathbf{r},\mathbf{y}} \quad \frac{1}{2} \cdot ||\mathbf{r}||_2^2 + \frac{\gamma}{2} \cdot ||\mathbf{X}\mathbf{r} - \mathbf{y}||_2^2$$

$$s.t. \quad \mathbf{y} \in \{0,1\}^{|\mathcal{C}|}, \; y_{i,j} = 1, \forall (u_i^{(1)}, u_j^{(2)}) \in \mathcal{A}^{(1,2)},$$

$$0 \leq \sum_{u_j^{(2)} \in G^{(2)}} y_{i,j} \leq 1, \forall u_i^{(1)} \in G^{(1)},$$

$$0 \leq \sum_{u_i^{(1)} \in G^{(1)}} y_{i,j} \leq 1, \forall u_j^{(2)} \in G^{(2)}$$

One-to-One Constraint

Learning Algorithm:

(1) Representation Learning Components Pre-training for latent feature representation learning, i.e., feature matrix **X**;

(2) Iterative optimization algorithm:

- Variable vector **r** inference
- Label vector **y** inference (to handle the one-to-one constraint)

Experimental Dataset

Experiment Dataset

TABLE I
PROPERTIES OF THE HETEROGENEOUS NETWORKS

		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	76,972
	write	9,490,707	48,756
	locate	615,515	48,756

 # Anchor Link: 3,388

Experimental Settings

Comparison Methods

	DETA	DETA_no	PALE [2]	DeepWalk+ TULER [3]+ Word2Vec (D+T+W)	DeepWalk [4]	Word2Vec [5]
Aligned Networks	√	√	√			
Heterogeneous Network	√	√		√		
Homogeneous Network			√		√	√

Metrics

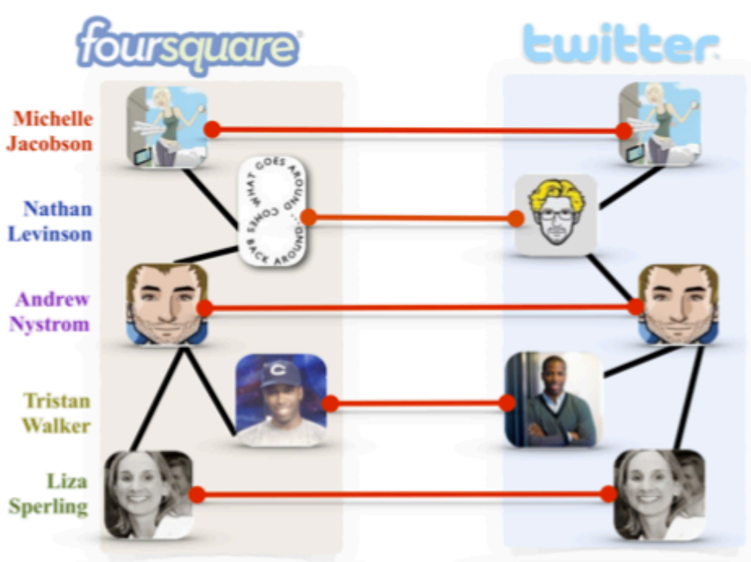
Accuracy	F1	Precision	Recall

[2] Man, Tong, Huawei Shen, Shenghua Liu, Xiaolong Jin, and Xueqi Cheng. Predict Anchor Links across Social Networks via an Embedding Approach. In IJCAI. 2016.
[3] Gao, Qiang, Fan Zhou, Kunpeng Zhang, Goce Trajcevski, Xucheng Luo, and Fengli Zhang. Identifying Human Mobility via Trajectory Embeddings. In IJCAI. 2017.
[4] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In KDD, 2014.
[5] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." In NIPS. 2013.

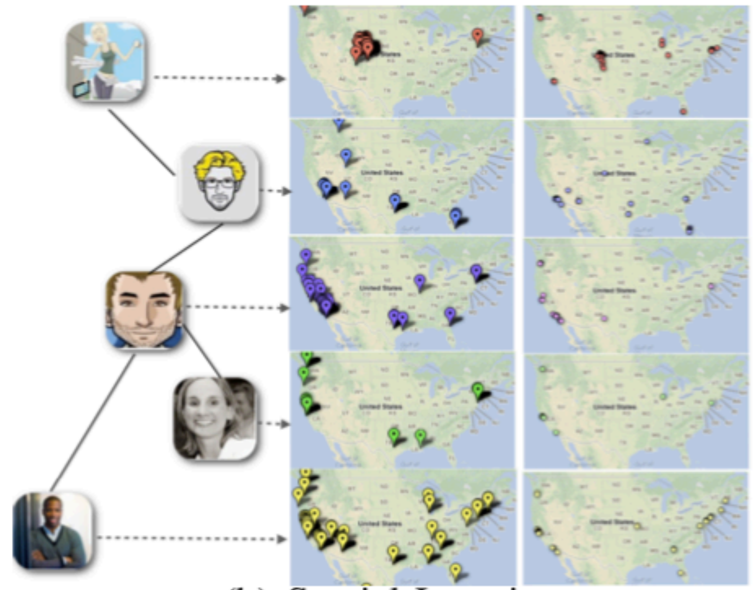
Experimental Result: Anchor Link Prediction

	method	Negative/Positive Ratio λ: 1	2	3	4	5	6	7	8	9	10
Accuracy	DETA	**0.924±0.016**	**0.904±0.008**	**0.901±0.009**	**0.903±0.008**	**0.902±0.008**	**0.919±0.004**	**0.954±0.020**	**0.965±0.003**	**0.965±0.004**	**0.967±0.003**
	DETA_NO	0.834±0.021	0.865±0.014	0.891±0.004	0.879±0.010	0.897±0.009	0.894±0.022	0.904± 0.021	0.926± 0.009	0.933± 0.007	0.933± 0.013
	PALE	0.517±0.022	0.370±0.016	0.316±0.012	0.290±0.007	0.281±0.011	0.280±0.007	0.285±0.001	0.294±0.005	0.305±0.004	0.315±0.008
	D+T+W	0.547±0.028	0.670±0.017	0.747±0.009	0.678±0.209	0.690±0.261	0.713±0.258	0.722±0.297	0.629±0.340	0.577±0.337	0.662±0.373
	DeepWalk	0.487±0.018	0.667±0.013	0750±0.009	0.800±0.009	0.833±0.008	0.857±0.005	0.875±0.005	0.889±0.005	0.900±0.003	0.910±0.005
	Word2Vec	0.484±0.013	0.667±0.016	0.750±0.009	0.800±0.009	0.833±0.007	0.857±0.005	0.875±0.005	0.889±0.005	0.900±0.003	0.909±0.005
F1	DETA	**0.928±0.016**	**0.870±0.009**	**0.829±0.014**	**0.797±0.014**	**0.760±0.024**	**0.743±0.008**	**0.813±0.050**	**0.824±0.015**	**0.798±0.024**	**0.791±0.025**
	DETA_NO	0.816±0.025	0.768±0.029	0.752±0.010	0.602±0.035	0.590±0.057	0.421±0.244	0.381±0.254	0.542±0.125	0.532±0.132	0.431±0.236
	PALE	0.667±0.020	0.498±0.019	0.402±0.014	0.334±0.012	0.283±0.012	0.247±0.010	0.223±0.009	0.200±0.012	0.182±0.007	0.166±0.011
	D+T+W	0.518±0.028	0.166±0.042	0.075±0.065	0.114±0.136	0.106±0.116	0.079±0.103	0.065±0.084	0.087±0.094	0.113±0.093	0.062±0.074
	DeepWalk	0.486±0.032	0.002±0.003	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000
	Word2Vec	0.326±0.326	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Precision	DETA	0.877±0.019	0.793±0.014	0.729±0.018	0.685±0.022	0.643±0.022	0.632±0.014	0.893±0.138	**0.965±0.010**	**0.963±0.010**	**0.962±0.037**
	DETA_NO	**0.913±0.017**	0.891±0.013	0.867±0.019	0.877±0.035	0.874±0.023	0.901±0.069	0.912±0.072	0.866±0.054	0.860±0.051	0.883±0.087
	PALE	0.509±0.021	0.339±0.017	0.258±0.011	0.206±0.009	0.170±0.009	0.145±0.007	0.129±0.006	0.115±0.007	0.103±0.004	0.094±0.007
	D+T+W	0.555±0.030	0.524±0.035	0.459±0.118	0.180±0.167	0.358±0.295	0.092±0.110	0.099±0.108	0.080±0.102	0.070±0.060	0.097±0.148
	DeepWalk	0.488±0.014	0.350±0.450	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000
	Word2Vec	0.242±0.243	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Recall	DETA	**0.986±0.016**	**0.965±0.014**	**0.961±0.021**	**0.954±0.014**	**0.930±0.035**	**0.901±0.012**	0.764±0.050	0.719±0.025	0.681±0.033	0.673±0.037
	DETA_NO	0.738±0.038	0.676±0.044	0.665±0.023	0.460±0.039	0.448±0.064	0.314±0.190	0.284±0.204	0.409±0.111	0.401±0.113	0.328±0.192
	PALE	0.966±0.009	0.937±0.012	0.922±0.013	0.891±0.014	0.854±0.020	0.829±0.022	**0.822±0.017**	**0.796±0.024**	**0.774±0.022**	**0.752±0.024**
	D+T+W	0.488±0.036	0.100±0.030	0.044±0.045	0.225±0.355	0.240±0.384	0.214±0.364	0.212±0.394	0.348±0.442	0.439±0.422	0.307±0.453
	DeepWalk	0.489±0.072	0.001±0.001	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000
	Word2Vec	0.500±0.500	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000

Experimental Result: Case Study



(a) Social Network



(b) Spatial Locations

user	word counts in both networks (Twitter, Foursquare)
Michelle Jacobson	art (65,2), style (16,3) audit (3,2), grill (19,2)
Nathan Levinson	happy (27,5), enjoy (9,4) week (18,4), shows (6,6)
Andrew Nystrom	awsm (2,3), kids (20,3) red (61,3), open (11,4)
Liza Sperling	ask (6,5), coffee (8,3) mochi (1,3), hangout (5,2)
Tristan Walker	win (19,4), amazing (55,5) awesome (51,4), please (9,4)

(c) Textual Word Usage

Fig. 4. Case study: five real-world users with their social, spatial and text distributions.

Foursquare \ Twitter	Tristan Walker	Nathan Levinson	Andrew Nystrom	Liza Sperling	Michelle Jacobson
Tristan Walker	1.559	1.637	1.764	1.825	1.999
Nathan Levinson	1.581	1.657	1.871	1.842	2.015
Andrew Nystrom	2.026	2.383	2.029	2.276	2.645
Liza Sperling	1.558	1.638	1.764	1.825	1.999
Michelle Jacobson	1.682	1.755	1.873	1.931	1.770

(a) Latent Social Feature based Distance Matrix

Foursquare \ Twitter	Tristan Walker	Nathan Levinson	Andrew Nystrom	Liza Sperling	Michelle Jacobson
Tristan Walker	198.489	216.454	338.386	257.475	268.829
Nathan Levinson	154.809	133.382	209.283	179.508	273.587
Andrew Nystrom	267.745	254.456	168.132	230.059	354.642
Liza Sperling	214.813	240.075	363.670	266.858	266.845
Michelle Jacobson	197.479	195.295	158.778	180.031	245.621

(b) Latent Trajectory Feature based Distance Matrix

Foursquare \ Twitter	Tristan Walker	Nathan Levinson	Andrew Nystrom	Liza Sperling	Michelle Jacobson
Tristan Walker	970.761	1164.241	1357.543	1068.010	33.790
Nathan Levinson	976.187	1159.558	1357.963	1071.939	38.973
Andrew Nystrom	987.637	1176.867	1370.252	1081.784	24.536
Liza Sperling	998.218	1189.149	1382.677	1088.606	14.737
Michelle Jacobson	999.859	1190.230	1384.597	1091.742	16.242

(c) Latent Text Feature based Distance Matrix

Fig. 5. Case Study Matrices

Broad Learning and Deep Social Network Alignment

Q&A

Contact Author: Jiawei Zhang
jiawei@ifmlab.org