SOCIAL MEDIA ANALYSIS

ACTOR NETWORK ANALYSIS & COOPERATION PREDICTION

G3 ｜朱柏諺、白日明、管晟宇、李家姍、蘇上青

June 2024

組內分工

組員	分工
R11725041 白日明 Jih-Ming Bai	文獻回顧、模型實驗
R12725028 管晟宇 Cheng-Yu Kuan	文獻回顧、Demo
B10704096 蘇上青 Shang-Qing Su	文獻回顧、資料收集
B10704031 朱柏諺 Po-Yen Chu	文獻回顧、EGES（建構模型、實驗）
B10704051 李家姍 Chia-Shan Li	文獻回顧、資料收集、簡報製作

AGENDA

1. Motivation
2. Literature Review
3. Dataset
4. Method & Experimental Results
5. Conclusion & Future Work

AGENDA

1	Motivation
2	Literature Review
3	Dataset
4	Method & Experimental Results
5	Conclusion & Future Work

The right cast determines the success

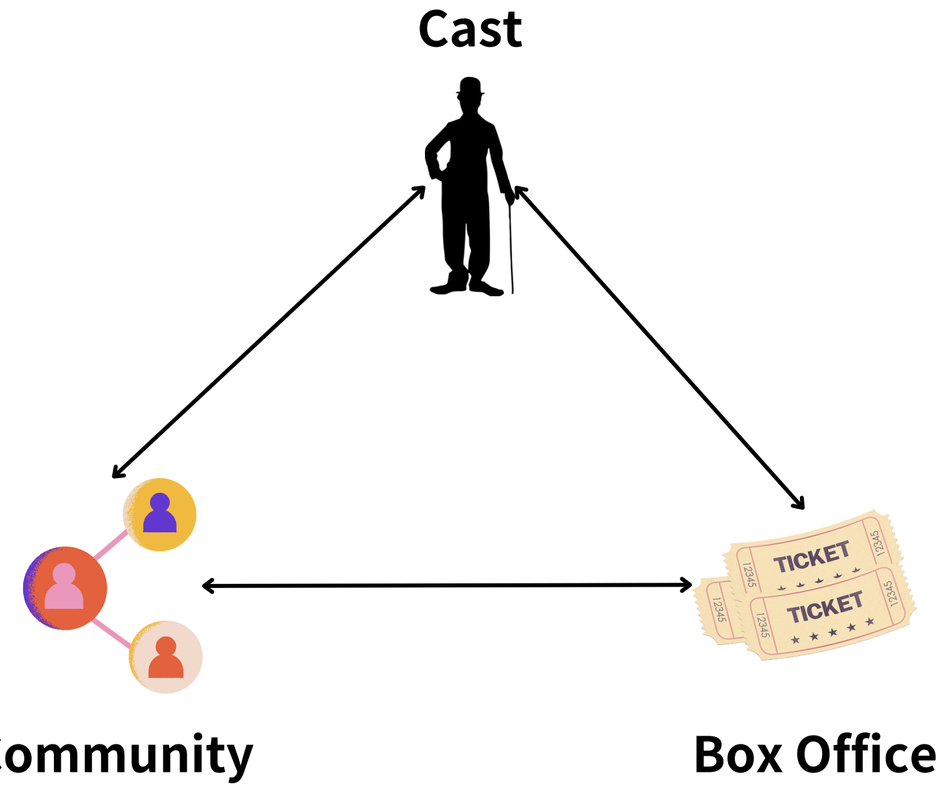
Success Example

Robert Downey Jr. and Chris Evans, portraying Iron Man and Captain America respectively in the Marvel Cinematic Universe, brought immense success to the franchise.

However, choosing the casts is difficult……

What is Difficult?

- Assessing the suitability of actors
- Evaluate whether the combination of actors would be liked by the audience
- Predicting the box office performance



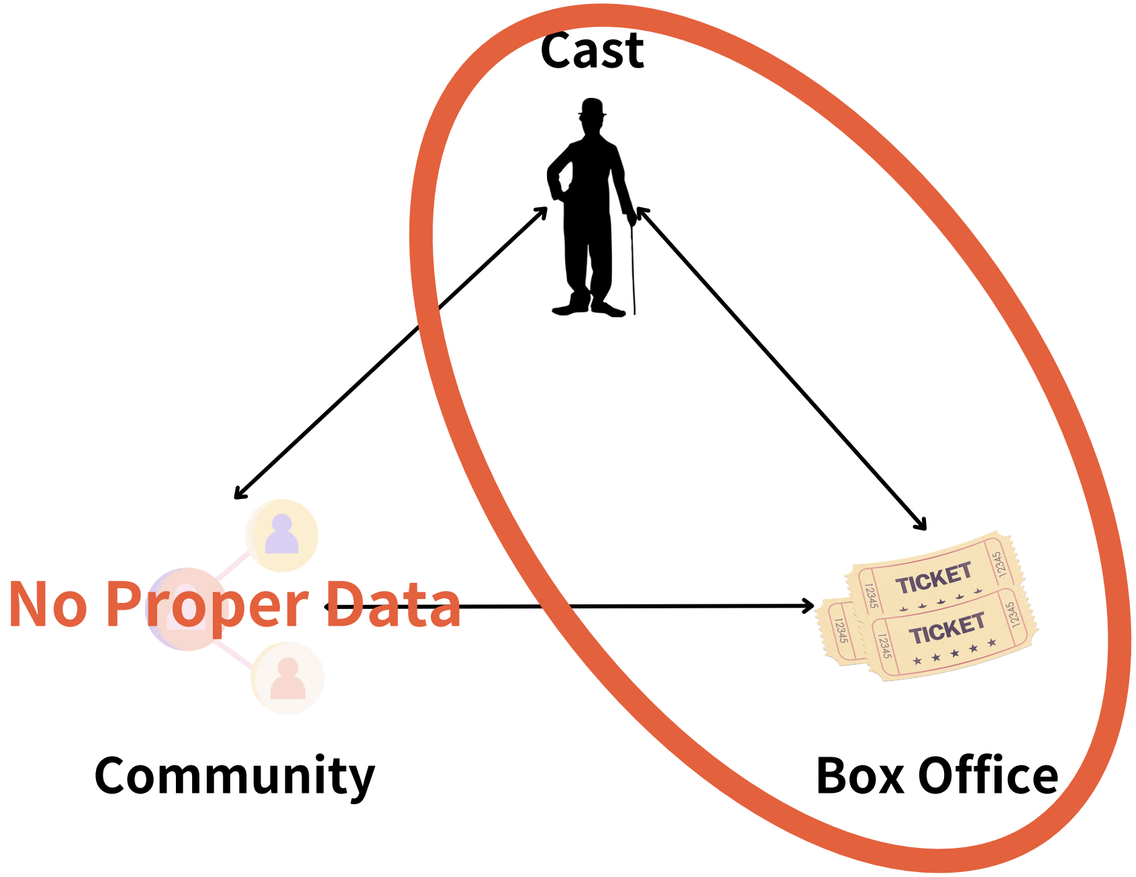
However, choosing the casts is difficult……

What is Difficult?

Assessing the suitability of actors

Evaluate whether the combination of actors would be liked by the audience

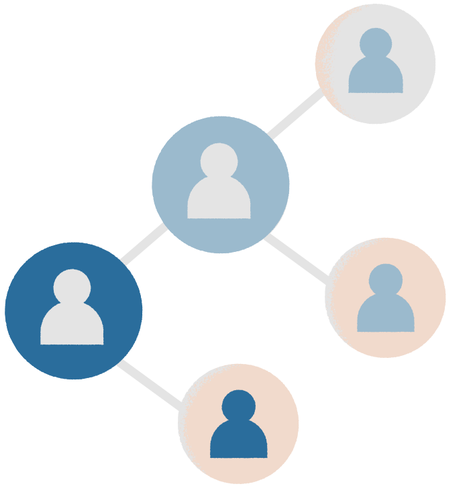
Predicting the box office performance



So, this study aims to……

Goal 1

analyze collaboration data among movie actors and construct networks of their collaboration relationships



Goal 2

predicting potential actor combinations and collaborative relationships

AGENDA

1. Motivation
2. Literature Review
3. Dataset
4. Method & Experimental Results
5. Conclusion & Future Work

Link Prediction Paradigms

Embedding

Preserves the edge structure of the graph.

- **EGES**
- metapath2vec, struc2vec, node2vec
- LINE, DeepWalk

Heuristic

- PPR, Graph distance

GNN

Encodes topological structures of graphs.

Autoencoder

- NESS
- VGAE

Subgraph

- BUDDY, ELPH
- **SEAL**
- GraphSAGE

Representation Learning Methods

	Training paradigm	Dataset	Release date
NESS	Autoencoder	Cora, Citeseer, PubMed (homophilous) WebKB, Chameleon (heterogeneous)	Mar 2023
BUDDY, ELPH	Subgraph	Cora, Citeseer, PubMed, OGB	Sep 2022
EGES	Embedding	Amazon Reviews, Taobao (private)	Mar 2018
SEAL	Subgraph	USAir, NS, PB, Yeast, C.ele, Power, Router, E.coli	Feb 2018
metapath2vec	Embedding	AMiner CS, DBIS	Aug 2017

AGENDA

1. Motivation
2. Literature Review
3. Dataset
4. Method & Experimental Results
5. Conclusion & Future Work

Dataset Overview

Movie Data	source	TMDB 15000 Movies Dataset (with credits) from Kaggle, Supplementary data from The Numbers website
	volume	2686 movies from 2000/01 to 2023/05
Actor Data	source	Movie_Actor Database from Data.World
	volume	1943 actors

How can we define a "successful" movie?

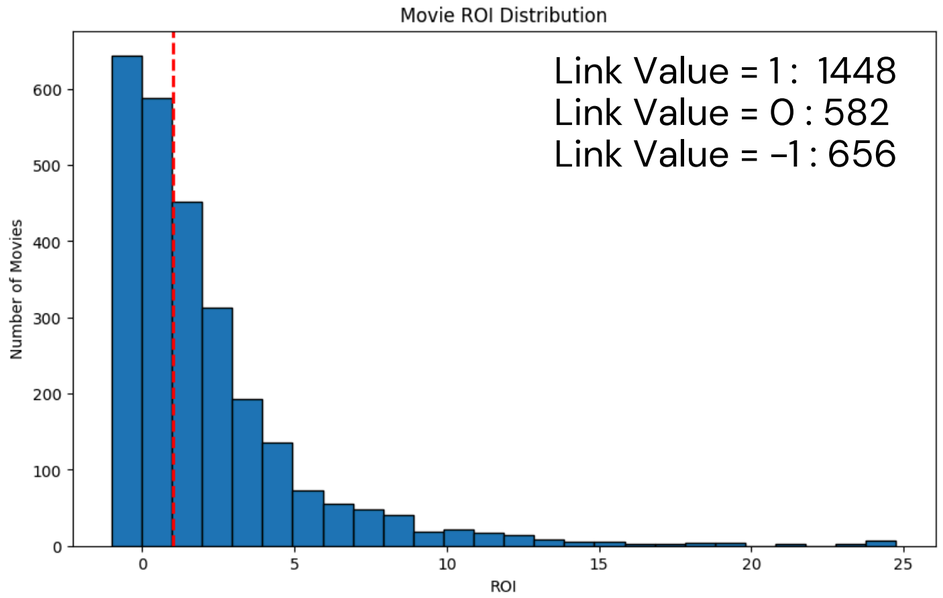
" A general rule of thumb for qualifying a movie as a “financial success” is to compare its gross revenue to twice its reported budget - in other words, an ROI of at least 1. "

ROI = **Gross - Budget** / **Budget**

ROI >= 1 -> Link Value = 1

1>ROI >= 0 -> Link Value = 0

ROI < 0 -> Link Value = -1



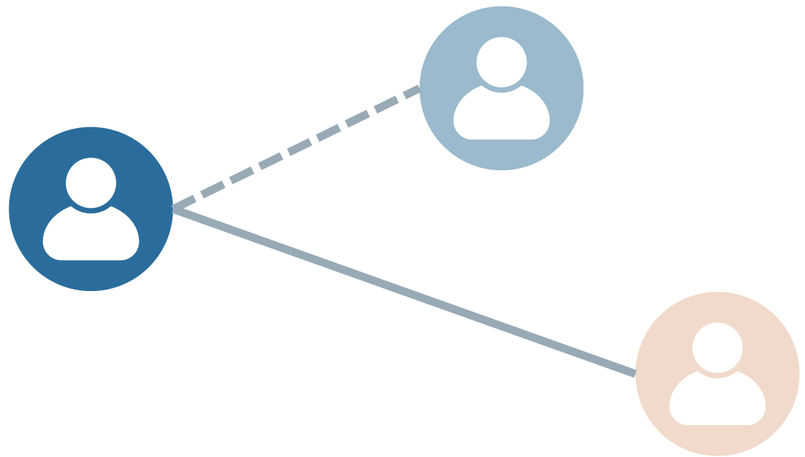
Actor_Network.json

Node


```
{
    "id": 0,
    "name": "AnneHathaway",
    "gender": 1.0,
    "popularity": "43.687",
    "Birth Country": "USA",
    "Height (Inches)": 68.0,
    "Ethnicity": "White",
    "NetWorth": 60000000.0,
    "Age": 43.0
}
```


Link


```
{
    "source": 0,
    "target": 2,
    "value": 1
 }
```

AGENDA

1. Motivation
2. Literature Review
3. Dataset
4. **Method & Experimental Results**
5. Conclusion & Future Work

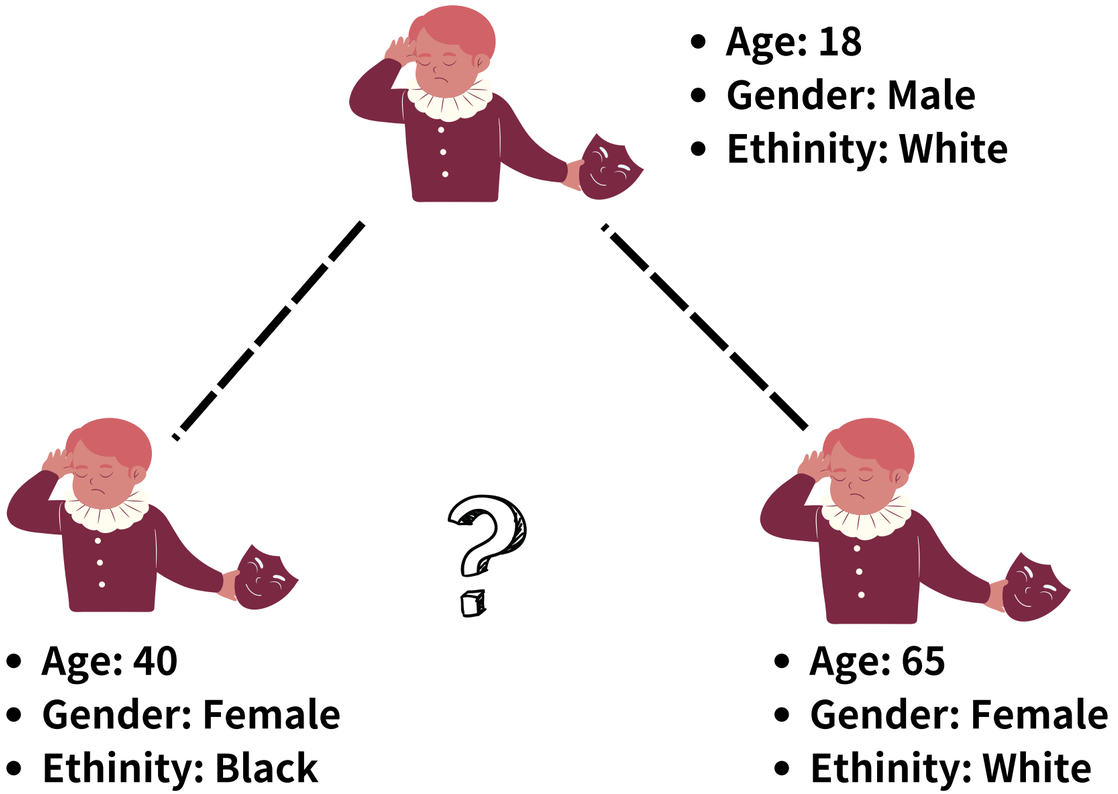
Link Prediction

predict the probability of success cooperation

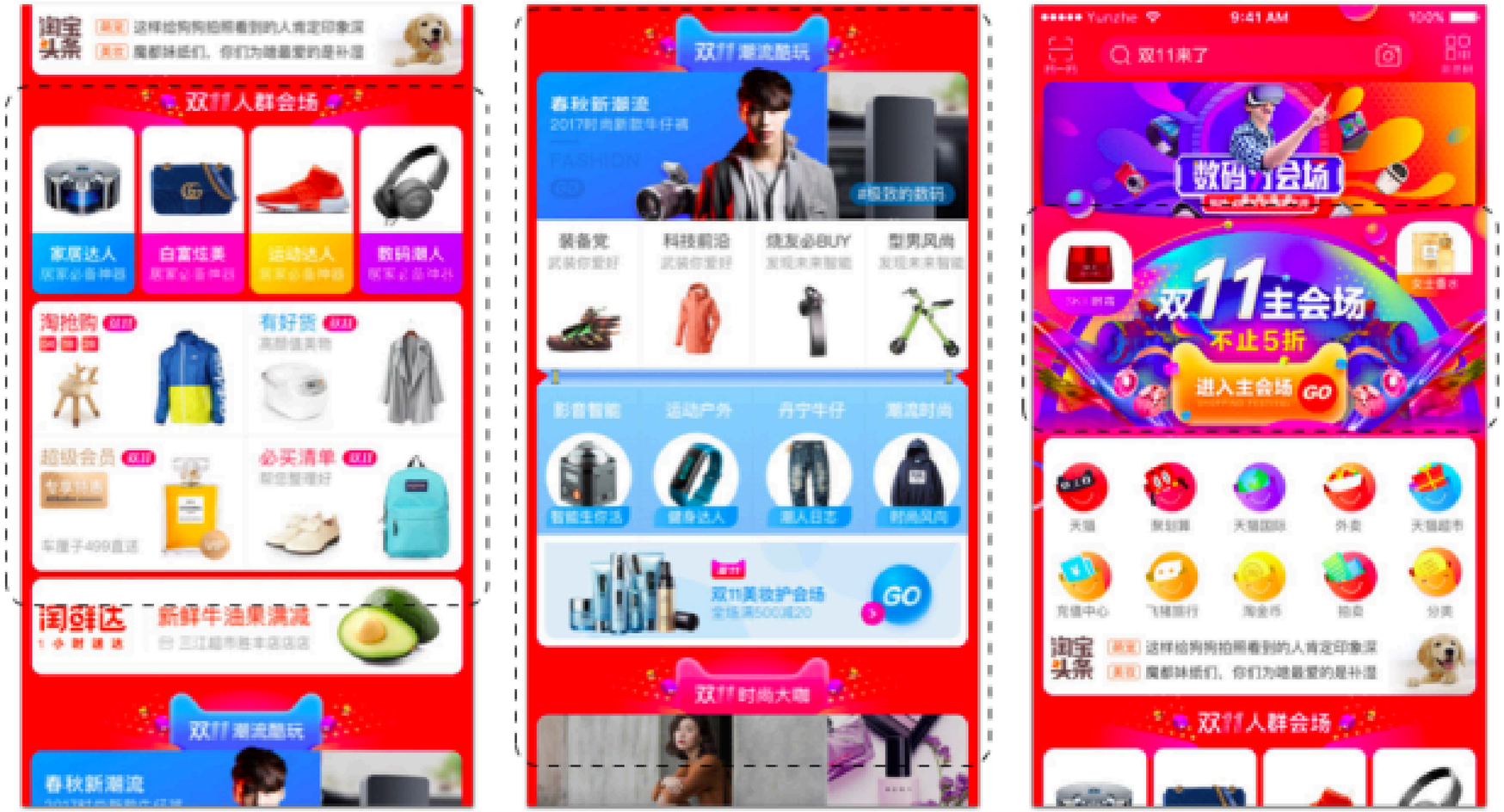
Link Prediction

EGES

SEAL



Node Embedding - EGES



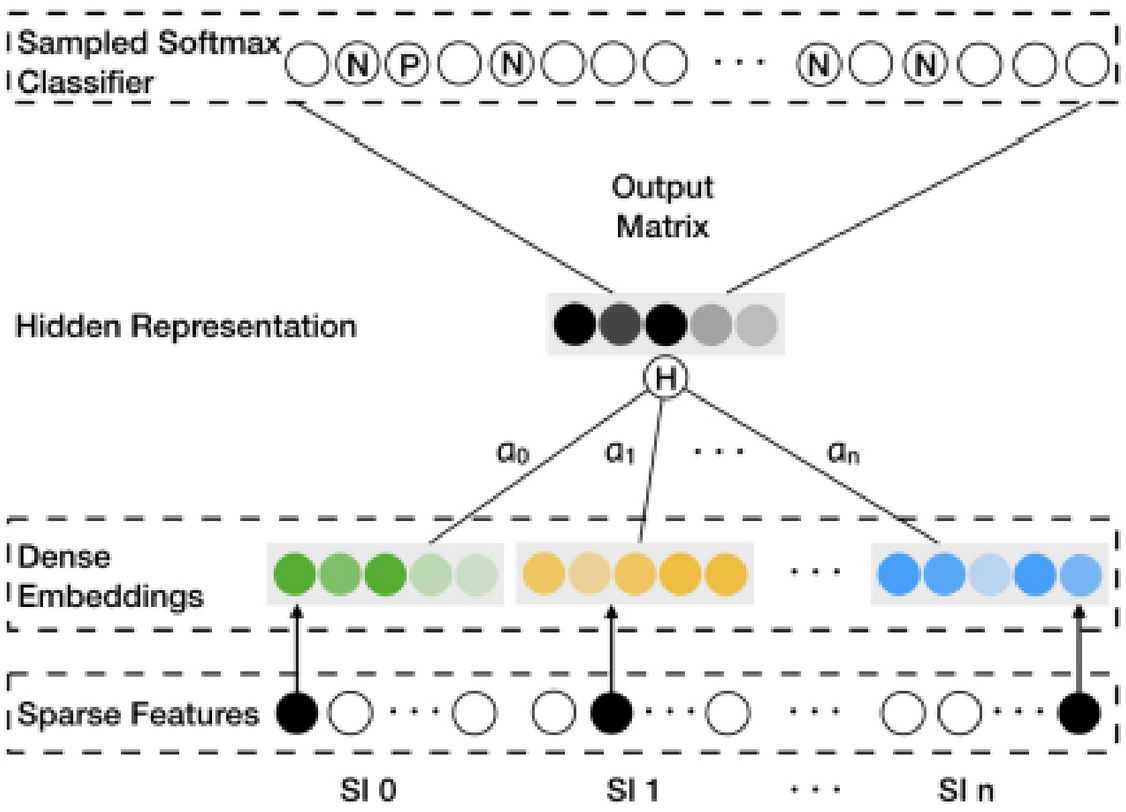
Reference: Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba

Node Embedding - EGES

Solve Cold-start issue

Add side information to embeddings

Enhanced = weighted



Reference: Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba

Node Embedding - EGES Implementation

Graph Definition	
Property	• directed -> undirected • weighted -> unweighted
Node	• actors
Other side information	• convert to categorical data
Link	• defined by popularity

Link Prediction Methods
XGBoost
Similarity Threshold
Similarity Ranking

Message Passing - GCN

Based on node embeddings, previous research often designed diverse message passing mechanisms to effectively consider other neighboring nodes in the network.

Initialize Node2Vec + Side Information

Information Propagation Graph Convolutional Network (GCN)

$$f^{gcn}(x_i) = \sum_{j \in N(i) \cup (j)} \frac{e_{j,i}}{\sqrt{\hat{d}_j \cdot \hat{d}_i}} x_j$$

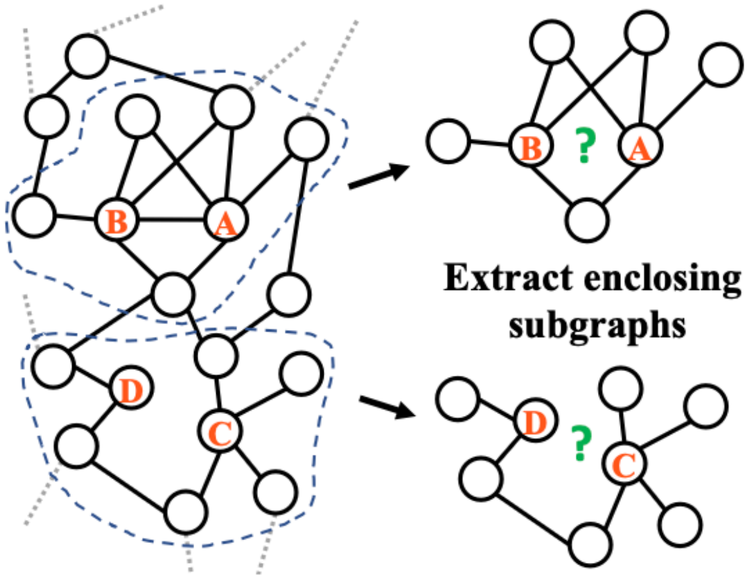
Prediction dense layer

Message Passing - SEAL

We implemented the SEAL (Subgraphs, Embeddings, and Attributes for Link prediction) algorithm proposed by Muhan Zhang and Yixin Chen, 2018.

It contains three steps:

1. **enclosing subgraph extraction**
   - Construct subgraph with nodes around target link
2. **node information matrix construction**
   - Node labeling (from subgraph) + node attributes + node embedding
3. **GNN learning**
   - GCN + dense layer



Experiment Results - Settings

Data Split

- Link Type: (+1), (0), (−1)
- Node Pairs w/o Link: (x)
- Split Ratio = 8:1:1

We utilize "Failure link(−1)" as negative samples in training set!

Train	Valid	Test
Pos: 1,051(+1) Neg: 651(-1)+400(x)	Pos: 131(+1) Neg: 131(x)	Pos: 131(+1) Neg: 131(x)

Evaluation Metrics : ROC-AUC

RQ1: How do our models perform?

- SEAL outperforms all the benchmark models.
- Models using graph information can perform better in this task.

	baseline	benchmark		best
	ML-based	**EGES**	**GCN**	**SEAL**
Valid	0.611	0.592	0.672	**0.845**
Test	0.553	0.592	0.676	**0.801**

RQ2: How does side information affect model performance?

- Actors' static features can help improve the collaboration recommendation

	Without Side Inforamtion			With Side Information		
Model	**EGES**	**GCN**	**SEAL**	**EGES**	**GCN**	**SEAL**
Valid	0.520	0.617	0.808	0.592	0.672	**0.845**
Test	0.565	0.609	0.770	0.592	0.676	**0.801**

RQ3: How does our model perform in application?

- Our model can accurately predict existing relationships

Cillian Murphy. & **Emily Blunt**

Oppenheimer (2023),
not in the dataset but is correctly predicted

RQ3: How does our model perform in application?

- Our model can accurately predict existing relationships

Emma Stone & **Mark Ruffalo**

Poor Things (2024),
not in the dataset but is correctly predicted

AGENDA

1	Motivation
2	Literature Review
3	Dataset
4	Method & Experimental Results
5	Conclusion & Future Work

Finally…

We used open source movie dataset to construct actor collaborative network, and achieved 80.1% ROC-AUC score by implementing GNN-based models.

Managerial Implication

We demonstrated a data-driven approach that allows online streaming platforms to find more suitable movie leads from a profitability perspective in film production.

Future Work

- More precise and directly relevant definition of “successful collaboration”
- Consideration of movie category and more comprehensive actor profile data
- Consideration of temporal relation and heterogeneous information

Thanks for Listening

PROPOSAL

References

Zhao, T., McAuley, J., & King, I. (2015, October). Improving latent factor models via personalized feature projection for one class recommendation. In Proceedings of the 24th ACM international on conference on information and knowledge management (pp. 821-830).

Chamberlain, B. P., Shirobokov, S., Rossi, E., Frasca, F., Markovich, T., Hammerla, N. Y., ... & Hansmire, M. (2022, September). Graph neural networks for link prediction with subgraph sketching. In The eleventh international conference on learning representations.

Zhang, M., & Chen, Y. (2018). Link prediction based on graph neural networks. Advances in neural information processing systems, 31.

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in neural information processing systems, 30.

Wang, J., Huang, P., Zhao, H., Zhang, Z., Zhao, B., & Lee, D. L. (2018, July). Billion-scale commodity embedding for e-commerce recommendation in alibaba. In Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 839-848).

Ucar, T. (2023). NESS: learning node embeddings from static subgraphs. arXiv preprint arXiv:2303.08958.

Kipf, T. N., & Welling, M. (2016). Variational graph auto-encoders. arXiv preprint arXiv:1611.07308.

組內分工

組員	分工
R11725041 白日明 Jih-Ming Bai	文獻回顧、模型實驗
R12725028 管晟宇 Cheng-Yu Kuan	文獻回顧、Demo
B10704096 蘇上青 Shang-Qing Su	文獻回顧、資料收集
B10704031 朱柏諺 Po-Yen Chu	文獻回顧、EGES（建構模型、實驗）
B10704051 李家姍 Chia-Shan Li	文獻回顧、資料收集、簡報製作