A Network Tour of the Football Transfer Market

"STUDY ON THE FOOTBALL PLAYER TRANSFER SYSTEM

IN EUROPE"

Saibo Geng

Olavo Bacelar

Xiao Zhou

Clarisse Poupon-Pourchot

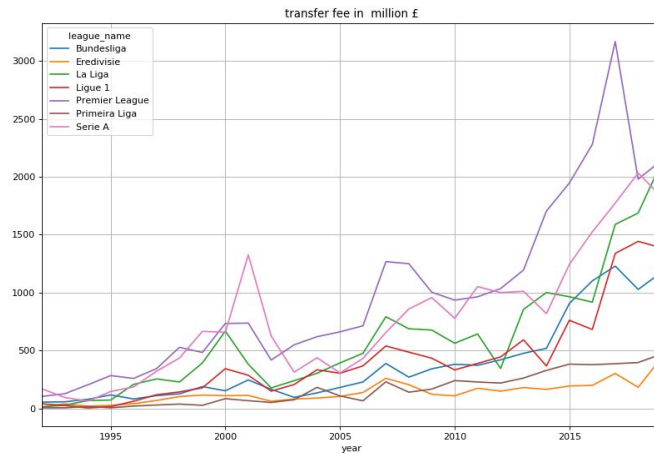**EPFL**

EE 558 A Network Tour of Data Science

Motivation

Transfer Market:
25.5 billion €

Research Question:
Investigate the network structure of football transfer market

transfer fee in million £

Overview of the project

A Network Tour of Football Transfer

- **0、FOOTBALL DATA COLLECTION**
  - European Football Transfers Dataset
  - Soccer Power Index
  - FIFA 2020 complete player dataset

- **1、EXPLORATION OF TRANSFERS**
  - DATA CLEANING
  - SIMPLE STATS ANALYSIS OF TRA
  - NETWORK STRUCTURE

- **2、TRANSFER MARKET ON GRAPH**
  - geographic distribution
  - Scale free VShomogeneous
  - transfer strategy and clustering coefficient
  - TOP 5 clubs in transfer market

- **3、CLUSTERS ON TRANSFER MARK**
  - K-means
  - modularity maximization

- **4、NEURALNETWORK WITHGRAPH**
  - features
  - performance and difficulty

Dataset

Kaggle Dataset

- European Football Transfers Dataset

- European soccer database (FIFA video games player attributes)

- FIFA 2020 Complete Database

Online Data

- Soccer power index data (FiveThirtyEight)

- Geographic coordinates of club stadium from Google

European Football Transfers Dataset

Original Dataset
(60420 transfers)
1992-2019

- Transfers of 9 leagues
- 11 features in this dataset
  - Club name
  - Player name
  - Transfer fee
  - etc

Not perfect: different names for the same club, same player

Solution: Fuzzy match + sometimes hardcode

Filtered Dataset (33423 transfers)

2000-2019

Selection of the clubs in the top 7 leagues:
- Premier League (ENG)
- Ligue 1 (FRA)
- Primeira Liga (POR)
- La Liga (ESP)
- Serie A (ITA)
- Bundesliga (DEU)
- Eredivisie (NLD)

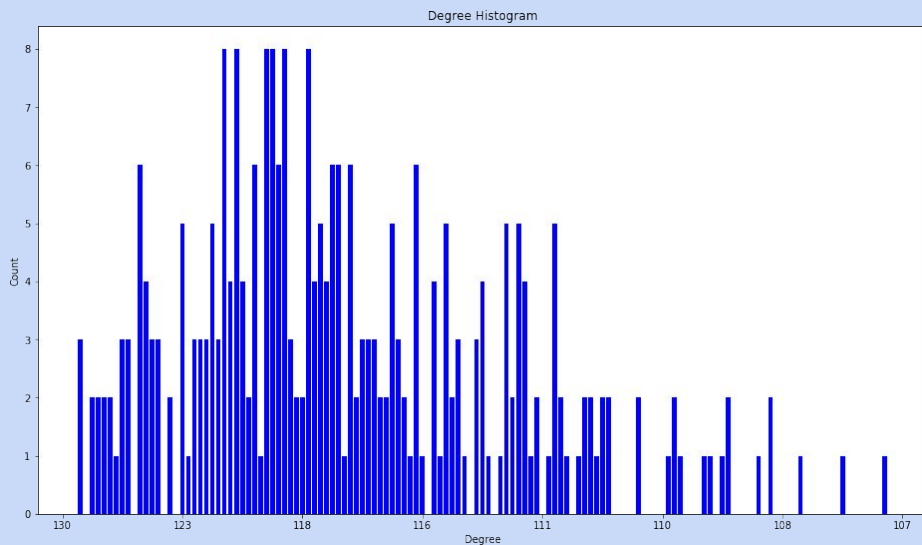# Created networks

Undirected networks, nodes = clubs

Different weight schemas:

1. Total number of transfers between pairs of clubs

2. Sum of transfer fees between pairs

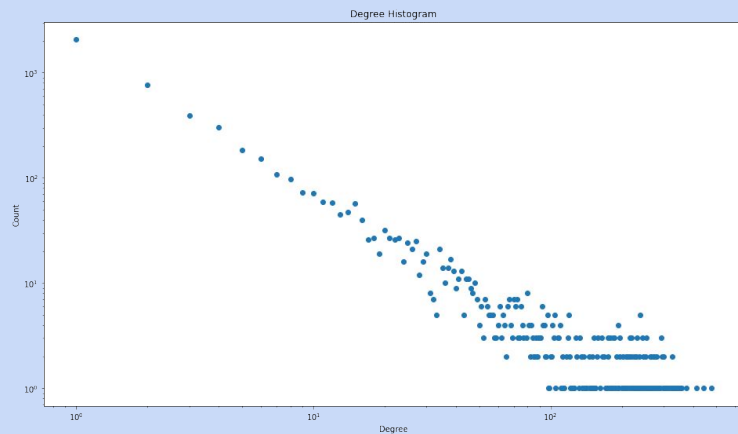3. (Harmonic mean of transfer fees between pairs)

graph signal
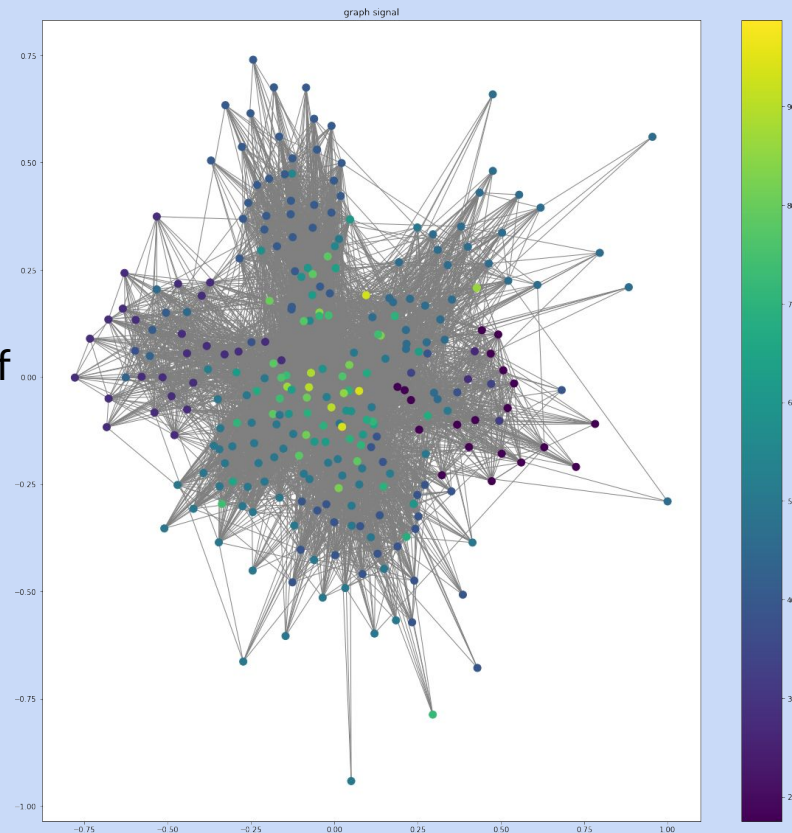
Histogram of total number of transfers

Transfers only between top 7 Leagues

All the transfers

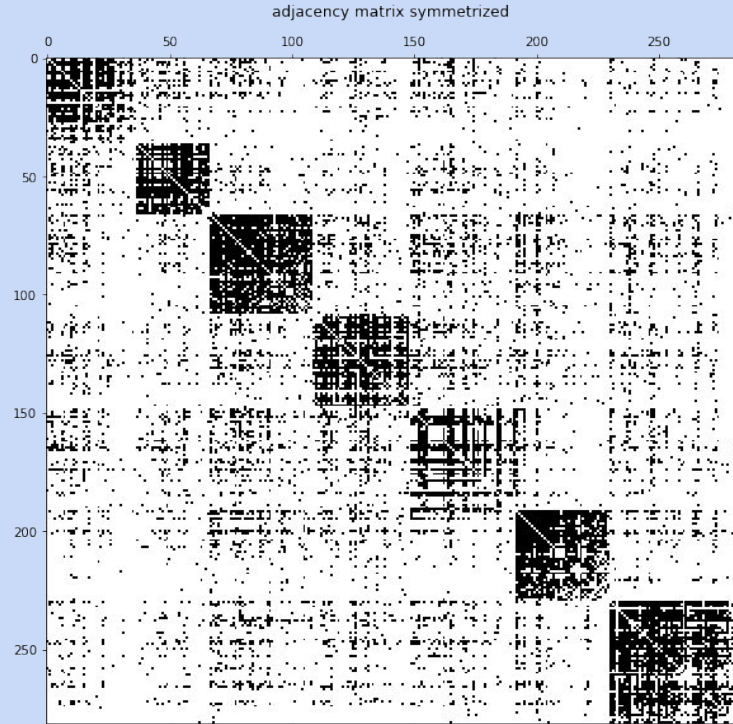# Graph Analysis - Adjacency Matrix

Number of nodes = 282

Number of edges = 6708

Average degree: 47.6
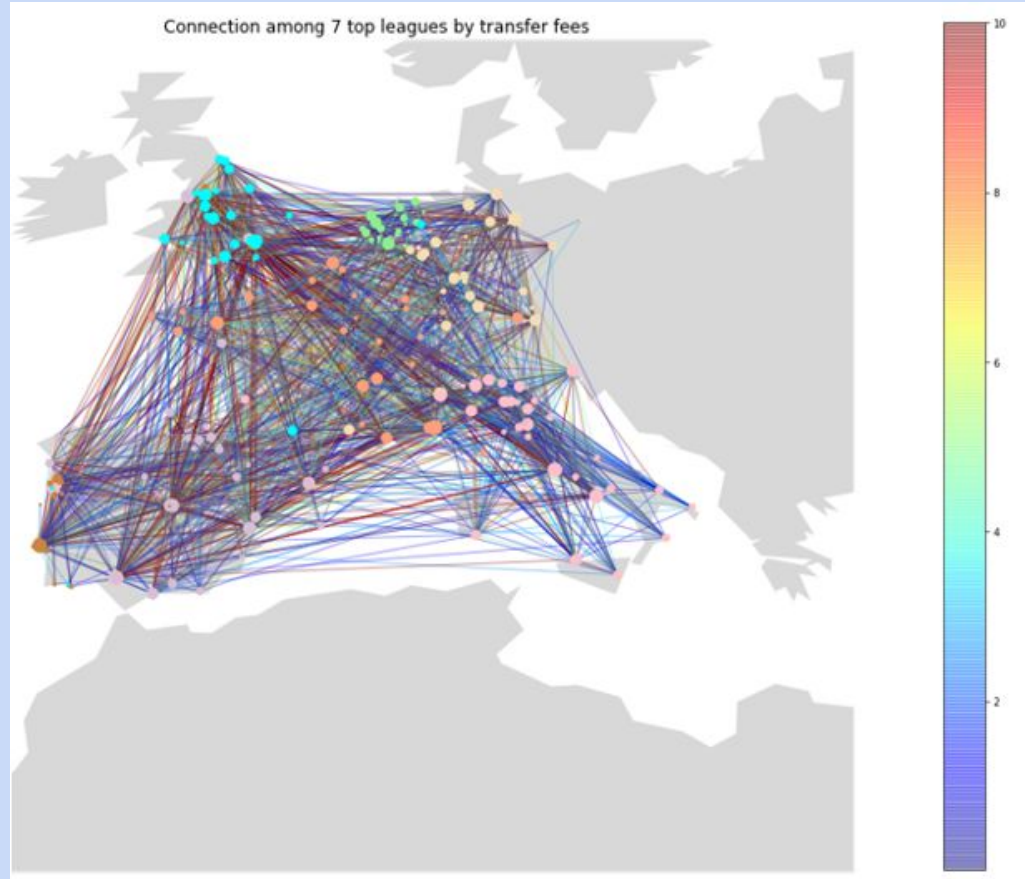
Average distance: 1.92

Diameter: 3

Sparsity: 0.17

adjacency matrix symmetrized

Connection among 7 top leagues by transfer fees

Geographical Visualization

transfer_fee_network

Transfer Fee Graph

i **More about this visualisation**

Legend:

● club_name

╲ fee_cleaned

◉ modularity

Search:

| Search by name | ▶ |
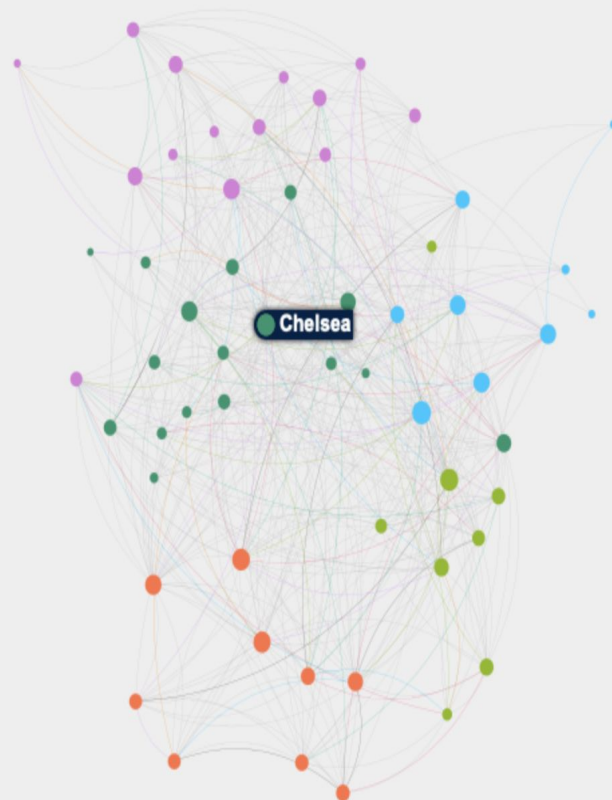
Information Pane

Chelsea

Modularity Class: 0

Eccentricity: 3.0

Degree: 57

Authority: 0.0074702455

Weighted In-Degree: 2.0

PageRank: 0.002748555588335911

Harmonic Closeness Centrality:
0.6358024691358022

Weighted Degree: 57.0

Strongly-Connected ID: 190
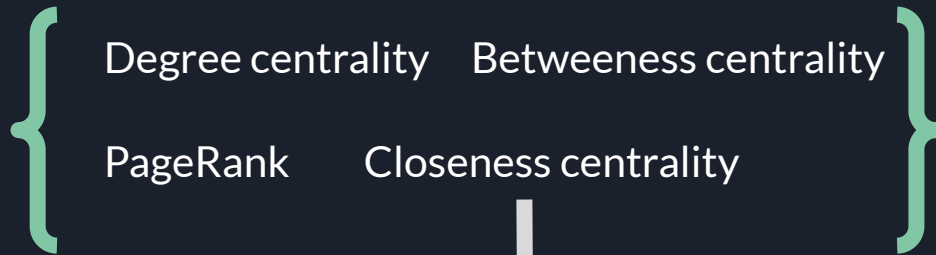
Weighted Out-Degree: 55.0

Out-Degree: 55

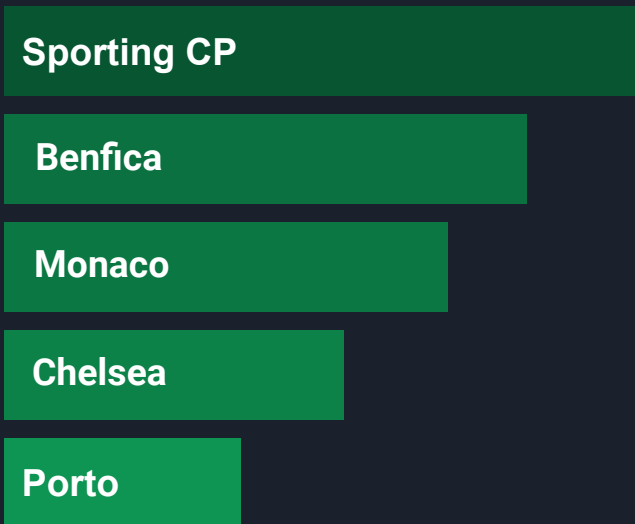**In-Degree:** 2

Closeness Centrality:
0.5658682634730539

Hub: 0.23109193

Component ID: 0

Eigenvector Centrality:
0.0038479375459430916

Chelsea

Mean of rank given by algorithms

Degree centrality   Betweeness centrality

PageRank   Closeness centrality

No fees:

Sporting CP

Benfica

Monaco

Chelsea

Porto

With transfer fee:

Porto

Liverpool

Roma

Benfica

Sevilla

Clustering Coefficient

Global clustering coefficient: 0.519

Most reputable clubs tend to have **lower** than average local clustering coefficient (0.35-0.48)

| club	Barcelona	Real Madrid	Paris Saint-Ger main	Juventus	Manchest er United	Bayern Munich
Clustering coefficient	0.58	0.45	0.47	0.44	0.39	0.47

COMMUNITY DETECTION

What could communities be in our network of transfer?

- Hypothesis
    - Communities can be a group of clubs having many transactions between them in the history.(partnership & privilege)
- Methods:
    - K-means Clustering
    - Modularity Maximization

COMMUNITY DETECTION-result

1.K-means clustering with K=7
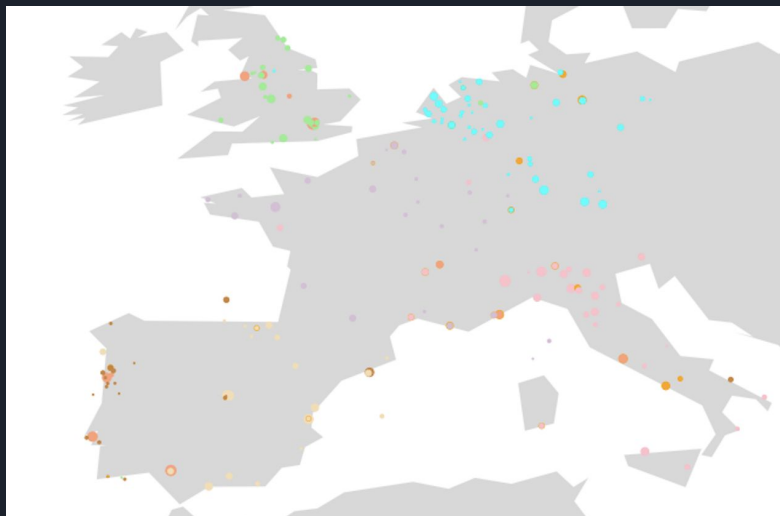
7 colors represent 7 clusters

- 🟢 UK
- 🔵 NLD+D EU
- 🌸 ITA
- 🟣 FRA
- 🟡 ESP
- 🟠 Other
- 🔴 PRT

2.K-means clustering with K=3

Cluster 0: Manchester City,Real Madrid, PSG, Barcelona,Ajax, Benfica + middle club from English League. Mean score=65

Cluster 1: middle clubs from French,German ,Spanish, Dutch league + Bayern Mean score=53

Cluster3:middle clubs from Italian ,French + Juventus Mean score= 48

K-means clustering with K=3

Cluster 0: Manchester City, Real Madrid, PSG, Barcelona, Ajax, Benfica + middle club from English League.

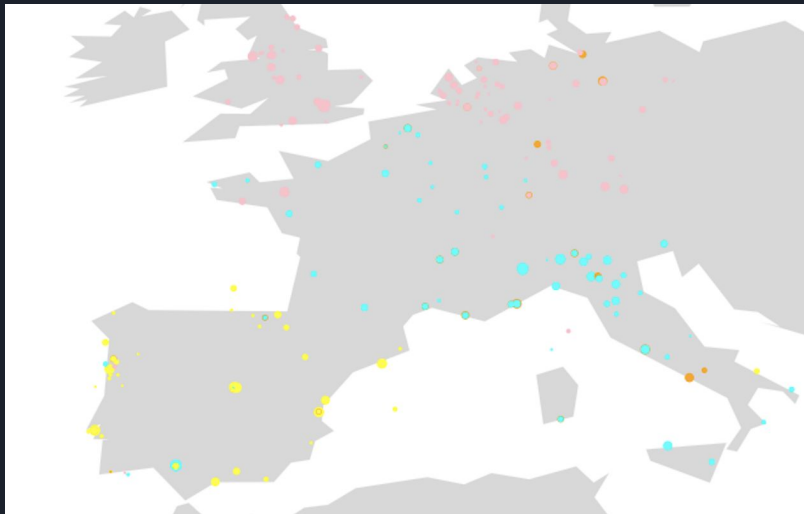**Mean score=65**

Cluster 1: middle clubs from French, German, Spanish, Dutch league + Bayern

Mean score=53

Cluster3: middle clubs from Italian, French + Juventus

Mean score= 48

1

2

3
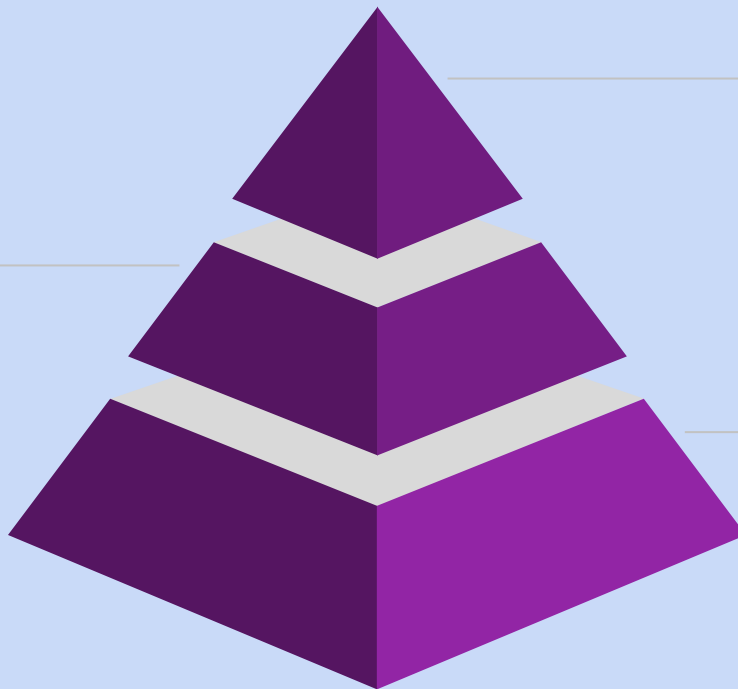
COMMUNITY DETECTION-result

2.Modularity Maximization(C=3)

● France+Italy

● Spain+Portugal

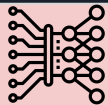● UK+Germany+
Netherlands

3 colors represent 3 clusters

Limits:

1. Hard clustering

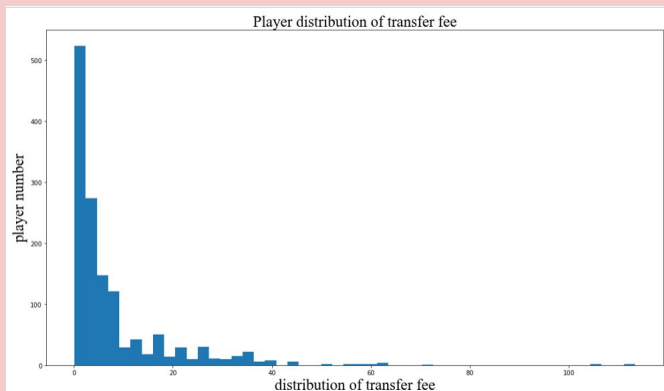2. High sparsity more sensible to noise

Significance:

1. Helps us to identify the leagues without requiring extra info.

2. Helps to identify strong clubs and divide clubs into categories by excellence
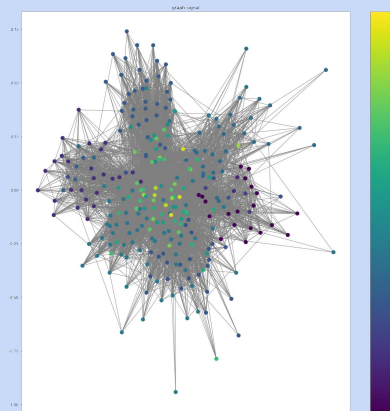
Neural Network

Goal:

Predict transfer fee class

for each transferred player
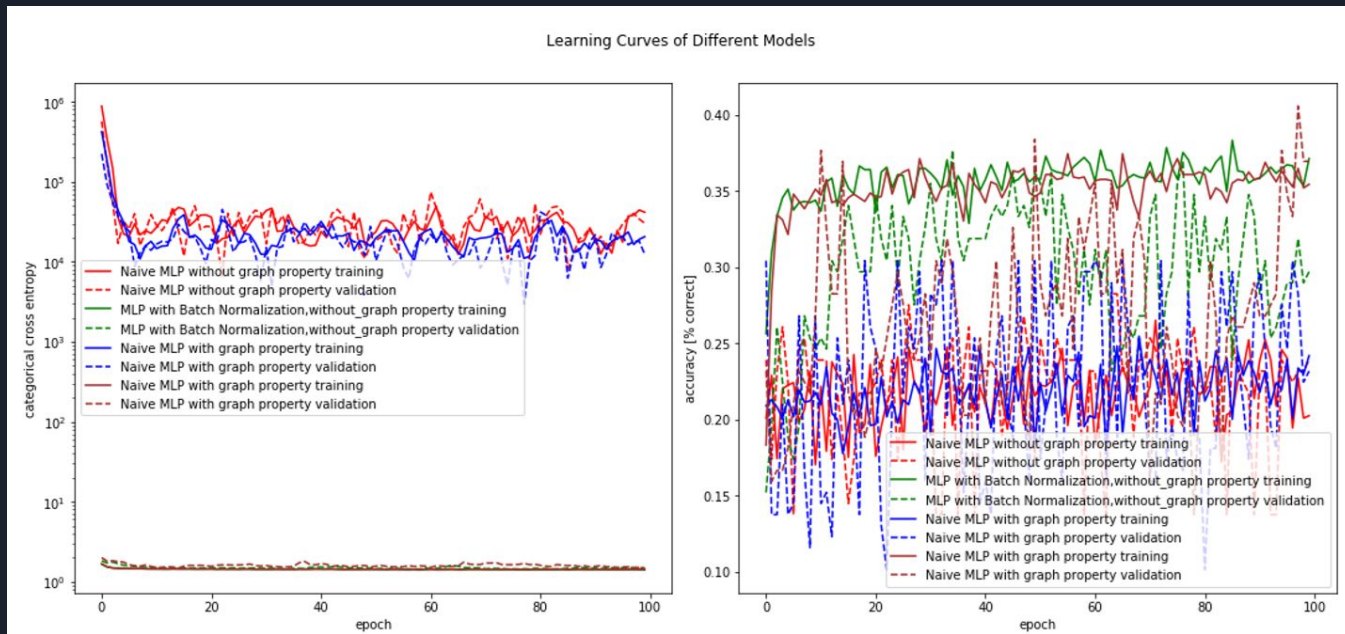
(unbiased categories)

Features

1. Transfer dataset features
2. FIFA game player attributes
3. Graph properties of club nodes

Neural Network

Predict transfer fee class with graph features

Model: 3-layer MLP with batch normalization

Conclusion

Transfer network analysis
- different facets of the economic importance of a club based on transfer records;

- go beyond economic or competitiveness measures;

Community detection
- recognize the affiliation relation of club and league from network;

Neural Network
- graph properties have a positive influence on the transfer fee prediction;
- consider time series model like LSTM in the future.

Thank you !