

Transfers & Language Network Analysis

Introduction

Data
Collection

Analysis &
Findings

Lessons
Learned

CATTELAN | LESSIS | WYSS

Analysis and Modelling of Social
Interactions | W.MSCIDS_AMS03.H2001



Research Question

Based on the assumption that Players transfer to regions with similar or same Language, can we identify patterns that contradict that assumption in the Swiss Football-League?

- **Data Collection**
- **Analysis**
- **Findings**

Transfers & Language Network Analysis



CATTELAN | LESSIS | WYSS

Analysis and Modelling of Social
Interactions | W.MSCIDS_AMS03.H2001

Data Needed:

- From Club --> To Club
- From Country --> To Country
- From Language --> To Language

Data Found:



Crawling (Chrauling)

Data Source
Quality

Execution



The Masterplan

Crawl Transfers



List of Clubs -->

Search Clubs



Club Profile -->

Crawl Country



The No-One-Cares-If-It-Works-Solution



Crawl Transfers

List of Clubs -->



Club Profile -->



Crawl Country

Search
"Transfer
markt" +
club +
"club
profile"

Countries

Languages

Google Is Sensible

- Use Multiple Agents
- Slowly

```
def country_crawler():
    global dict_countries
    try:
        for team in all_teams_unique[len(dict_countries):]:
            query = "transfermarkt club profile " + team
            for j in search(query, tld="co.in", num=1, stop=1, pause=60):
                headers2_agent = random.choice(headers2_list)
                headers2 = {'User-Agent': headers2_agent}
                pageTree = requests.get(j, headers=headers2)
                soup_country = BeautifulSoup(pageTree.content, 'html.parser')
                try:
                    country = soup_country.find('div', id ="verein_head").find('span','mediumpunkt').find('img')
                    dict_countries[team] = country
                except Exception:
                    dict_countries[team] = 'Not Found'
                print(len(dict_countries))
            with open('dict_countries_output.csv', 'w', encoding = 'utf-8') as csv_file:
                writer = csv.writer(csv_file)
                for key, value in dict_countries.items():
                    writer.writerow([key, value])

        except Exception:
            print("trying to work around connection error")
            time.sleep(120)
            country_crawler()
```

Languages

manually

~90 Countries



Season	Name	Transfer Sum	From	FromCountry	FromLanguage	FromLanguageFamily	To	ToCountry	ToLanguage	ToLanguageFamily
20/21	Arthur Cabral	4,40 Mio. €	Palmeiras	Brazil	Portuguese	Romance	FC Basel	Switzerland	German	Germanic
20/21	Kaly Sene	4,00 Mio. €	Juventus U19	Italy	Italian	Romance	FC Basel	Switzerland	German	Germanic
20/21	Edon Zhegrovci	3,00 Mio. €	KRC Genk	Belgium	French	Romance	FC Basel	Switzerland	German	Germanic
20/21	Andrea Padula	450 Tsd. €	FC Wil 1900	Switzerland	German	Germanic	FC Basel	Switzerland	German	Germanic
20/21	Heinz Lindner	abkömmlingfrei	Wehen Wiesbaden	Germany	German	Germanic	FC Basel	Switzerland	German	Germanic
...
11/12	Mickaël Facchinetti	Leihe	FC Lugano	Switzerland	Italian	Romance	Chievo Verona	Italy	Italian	Romance
11/12	Palmo Di Dio	-	FC Lugano	Switzerland	Italian	Romance	FC Lugano U21	Switzerland	Italian	Romance

Transfers & Language Network Analysis

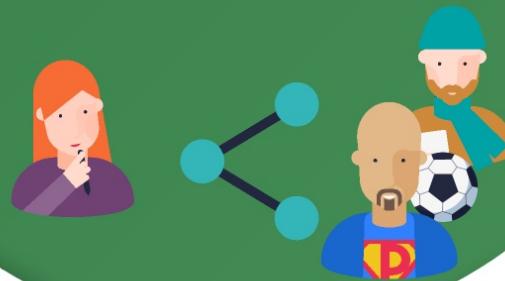


CATTELAN | LESSIS | WYSS

Analysis and Modelling of Social
Interactions | W.MSCIDS_AMS03.H2001

Networks

- Data preparation
- Key metrics
- Network structure
- Community detection & analysis



Prep

Analysis

Findings

Data understanding & preparation

WORK WORK WORK
85 % Python | 5 % tableau | 10 % Gephi



Gephi

Graph Metrics

1. Communities

Clauset-Newman-Moore greedy algorithm
multiple connections **45 communities**
single connection **13 communities**

Type: **MultiDiGraph**

Number of nodes: 900
Number of edges: 4028
Average in degree: 4.4756
Average out degree: 4.4756

2. Modularity

overall modularity is around **0.26**
low, but no impact on network or ability to exchange players

Type: **MultiGraph**

Number of nodes: 900
Number of edges: 4028
Average degree: 8.9511

3. Degree Centrality (important)

['FC Lugano', 'FC Sion', 'Grasshopper Club Zürich', 'FC Basel']

Type: **Graph**

Number of nodes: 900
Number of edges: 1659
Average degree: 3.6867

4. Betweenness Centrality (not important)

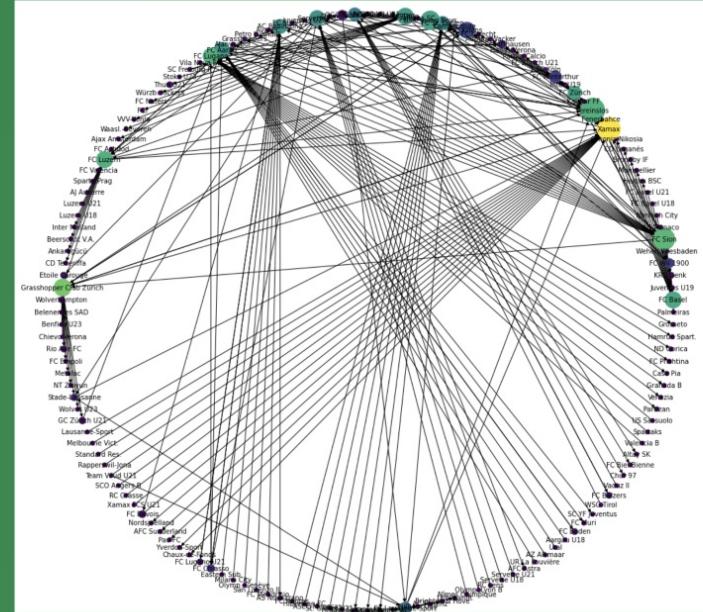
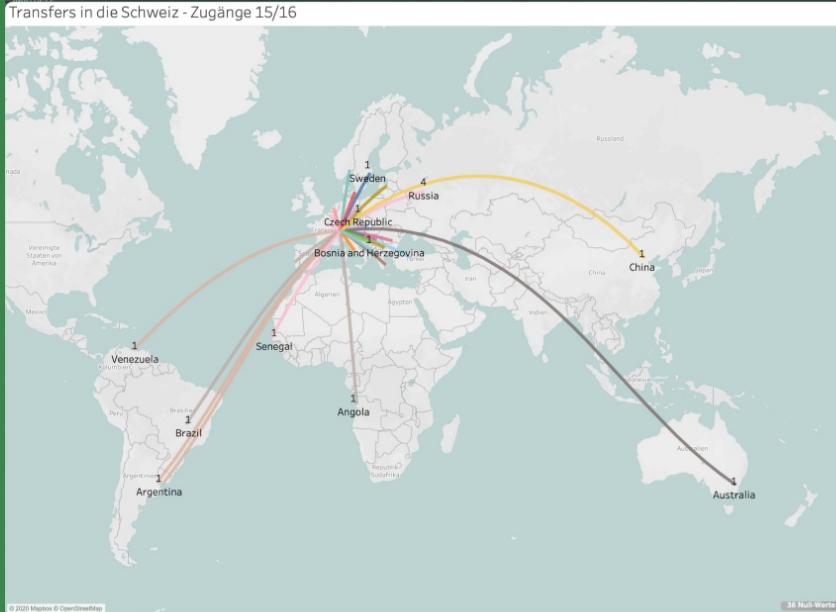
['FC Lugano', 'FC Sion', 'FC Lausanne-Sport', 'Grasshopper Club Zürich']

average clustering: 0.28
(young players from small clubs)

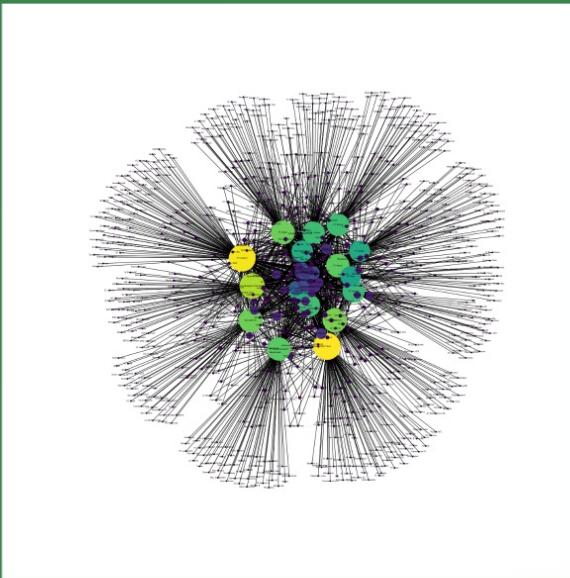
5. Eigenvector Centrality (not on multi / directed applicable)

['FC Sion', 'FC Basel', 'FC Lugano', 'FC St. Gallen']

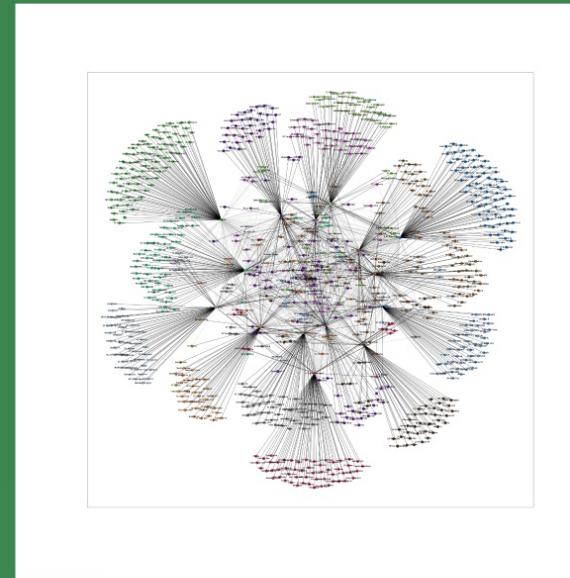
Where did he go, where does he come from?



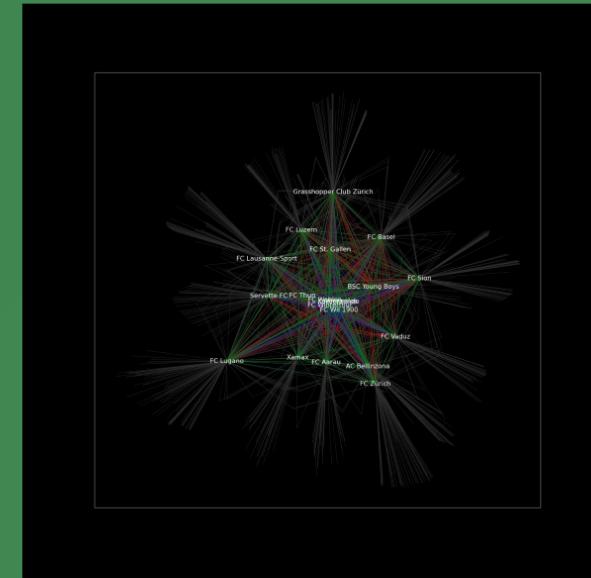
Visual Analysis



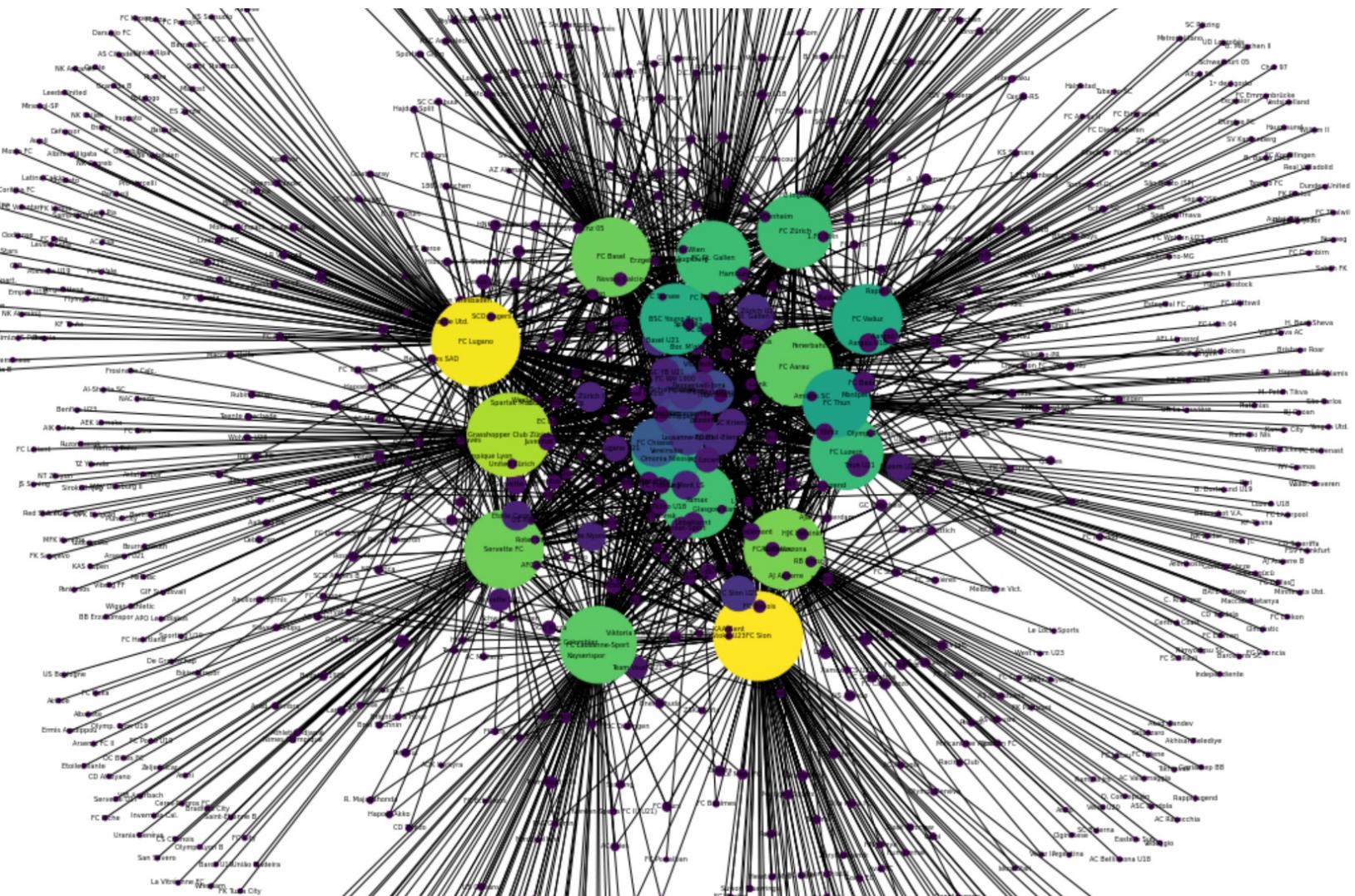
DEGREE CENTRALITY

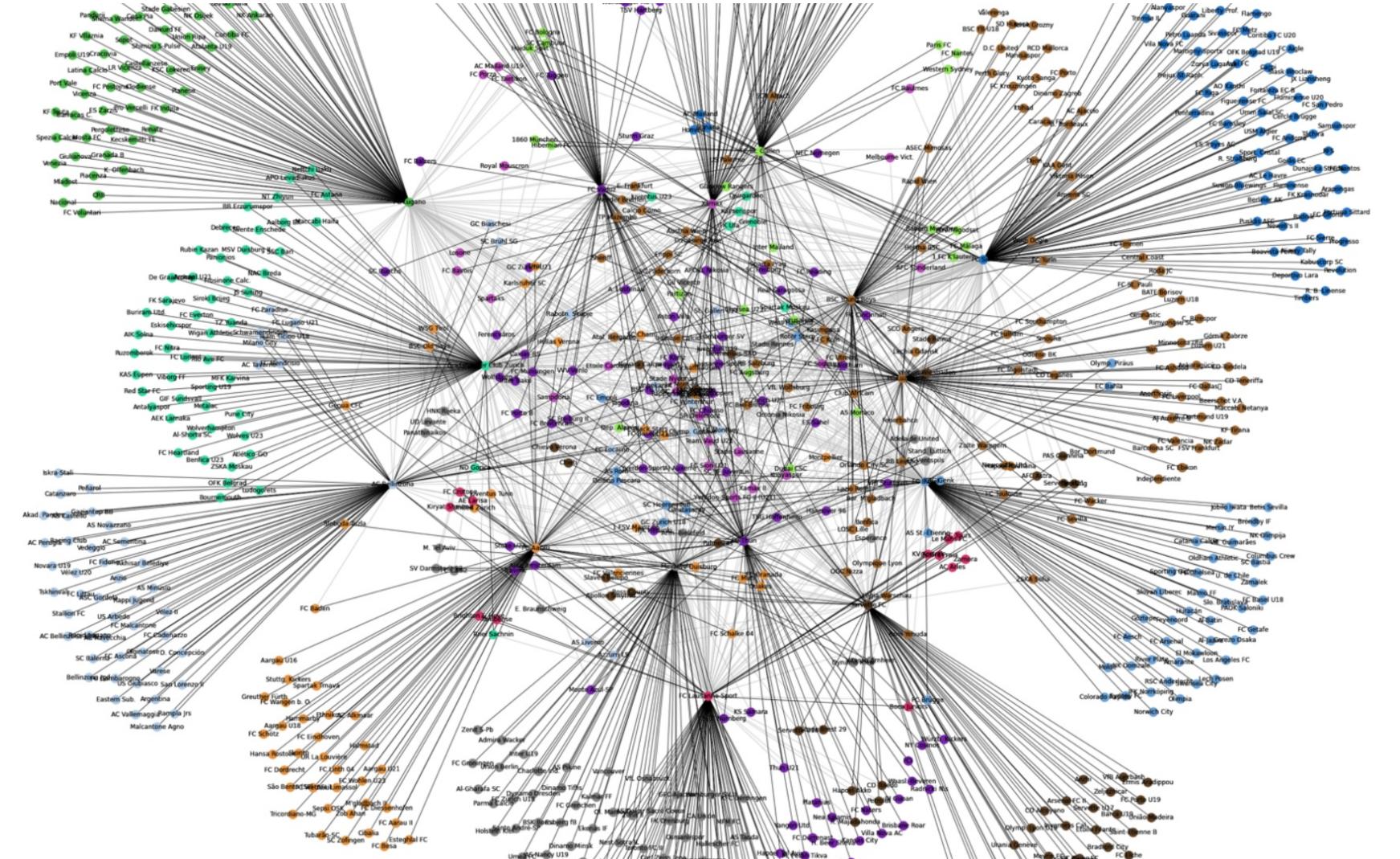


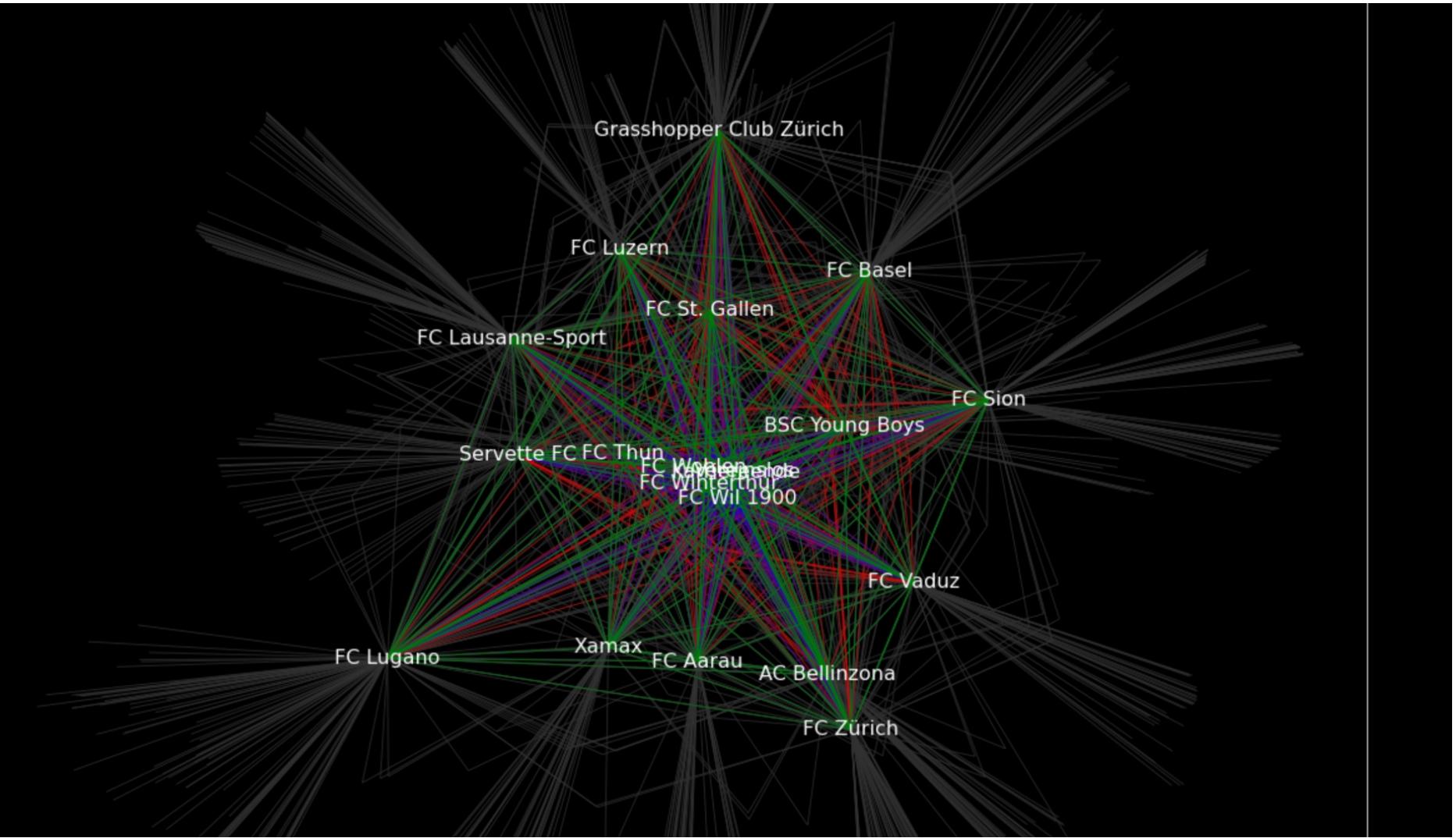
PROBABILISTIC COMMUNITY



K Core ($k = 15$)





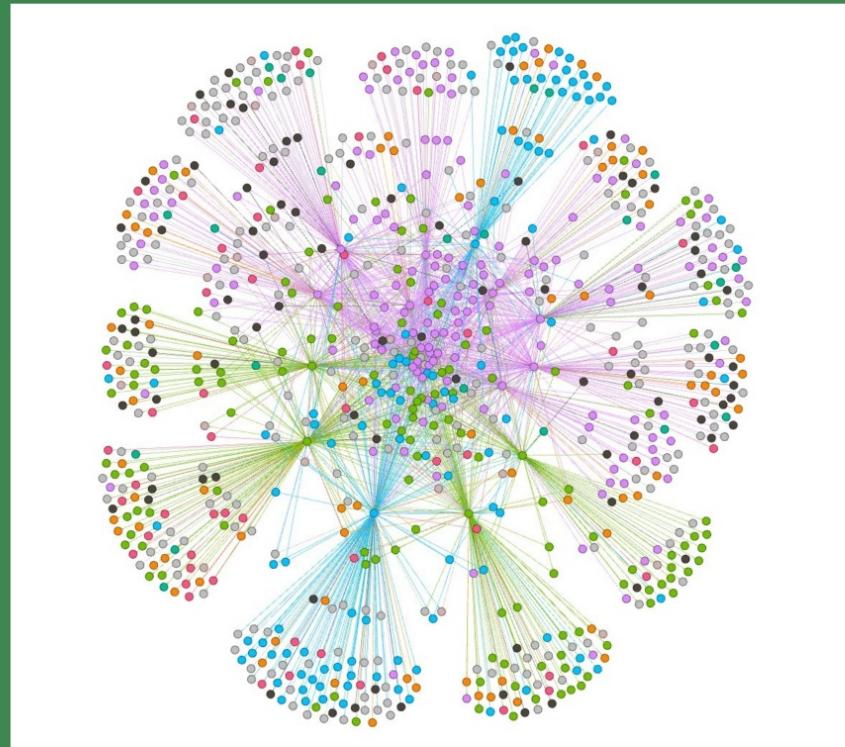


Transfer by language

Nodes = clubs
Edges = transfers

German	(17.48%)
French	(15.82%)
Italian	(12.39%)
English	(8.85%)
Spanish	(8.3%)
Portuguese	(6.08%)
Turkish	(2.54%)
Greek	(2.21%)
Dutch	(2.21%)
Russian	(1.88%)
Arabic	(1.66%)
Serbian	(1.66%)
Hebrew	(1.55%)

Strong mix within groups.
Many clubs still transfer
irrespective of their own
language and origin.



Yes but No

YES

We found out that surprisingly **FC Lugano** is the **most important** club regarding the swiss player transfer network
(degree centrality)

FC Sion regarding its vip connections
(eigenvector centrality)

We found out that **clubs** most likely transfer a player which speaks the **same language** as spoken at the clubs hometown but still transfer very mixed.

NO

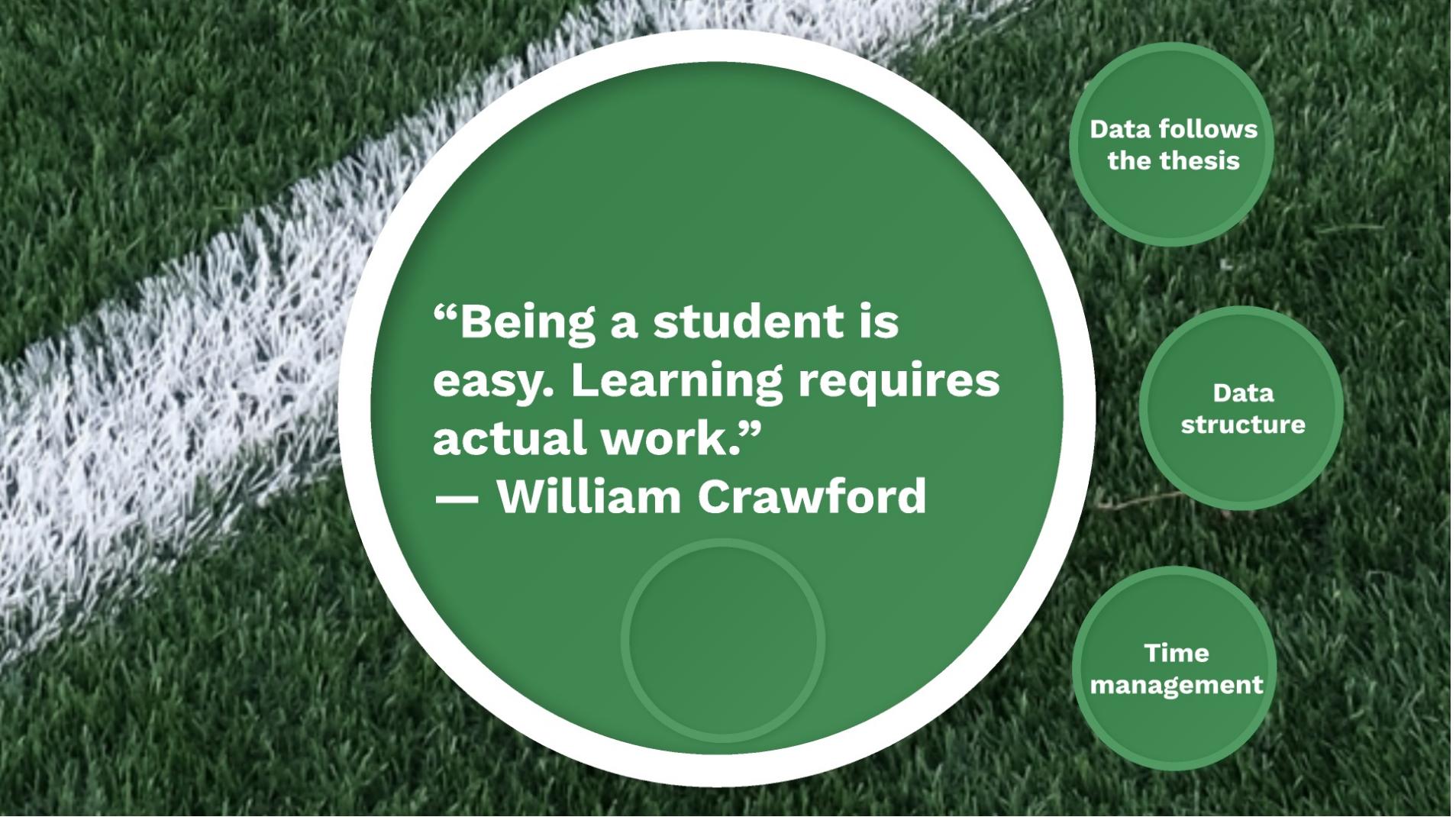
We have not found irregularities within language groups (there might be)

Transfers & Language Network Analysis



CATTELAN | LESSIS | WYSS

Analysis and Modelling of Social
Interactions | W.MSCIDS_AMS03.H2001



**“Being a student is
easy. Learning requires
actual work.”**

— William Crawford

Data follows
the thesis

Data
structure

Time
management

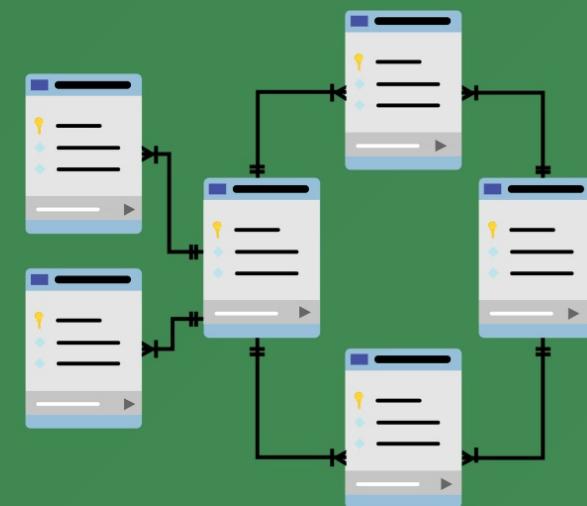
Thesis defines data needed

- Thesis as a starting point
- Data adapted to thesis



Adapting data-source on the way

- Data not perfect for analysing the thesis
- Scrape larger parts of transfermarkt.ch
- More seasons



Time is running



Web scraping is very time-consuming. We spent several hours scraping when we could have used it for analysis. #allnighters
A solid timetable could help staying on track.

Thank you!



SCAN TO OPEN GITHUB REPO



LAUGH TO FORGET CORONA
(at least for a moment)

Transfers & Language Network Analysis



CATTELAN | LESSIS | WYSS

Analysis and Modelling of Social
Interactions | W.MSCIDS_AMS03.H2001