

NETWORK ECONOMY REPORT

PREPARED BY

DAVIDE LOCCI
ALESSANDRO PIRODDI
ALICE PORTA



Report progetto

Economia dei network

Football Market Network Analysis on top European
Leagues.

Focus on the Italy case.



A.a. 2021/2022

Corso di studi: Economia dei network

Prof. Stefano Matta

Davide Locci 11/82/00197

Alessandro Piroddi 11/82/00205

Alice Porta 11/82/00142

Indice

Introduzione

Quesiti posti

Dataset

Descrizione

Pre-processing

Analisi esplorativa

Parte core

Primo punto

Secondo punto

Terzo punto

Quarto punto

Conclusioni

Bibliografia

INTRODUZIONE

Il presente lavoro si pone come obiettivo l'analisi, tramite network, dei trasferimenti avvenuti in entrata ed in uscita, nelle squadre delle sette maggiori leghe europee in un arco temporale che comprende gli ultimi 15 anni, dal 2007 al 2021.

Le sette leghe prese in considerazione sono rispettivamente: Serie A italiana, Premier League inglese, Liga spagnola, Bundesliga tedesca, Ligue1 Francese, Eredivisie olandese, Primeira liga portoghese.

Tutte e sette rappresentano il top level di ciascuna nazione presa in considerazione.

Alle leghe citate appena sopra si aggiungono le serie minori riguardanti il campionato italiano.

Infatti nell'analisi, pur prendendo sporadicamente in considerazione anche i campionati stranieri (soprattutto come metro di paragone), l'obiettivo è quello di operare un focus che vada a fondo sui trasferimenti del calcio italiano. Per questo abbiamo deciso di tenere in considerazione, per l'Italia, non solo la Serie A ma anche le leghe minori professionalistiche Serie B e Serie C, spingendoci anche alla lega dilettantistica Serie D, ed al campionato Primavera dei giovani.

L'analisi si articola in una parte introduttiva ed in una parte di analisi vera e propria attraverso i grafici.

Parte introduttiva: La prima parte consiste in una panoramica generale, per fornire, tramite grafici commentati, un primo impatto di quelli che sono i grandi numeri che caratterizzano il calciomercato europeo ed italiano.

Parte core: La parte core invece si serve come strumento primario dei grafici, supportati da ulteriori grafici per mostrare le misure prese in considerazione. Essa si articola a sua volta in 4 punti.

Il lavoro svolto è visionabile ai seguenti link:

Github 



Google colaboratory 



Le Leghe



QUESITI POSTI

Il primo punto 1a è l'unico in cui il primo grafo non è preceduto da una domanda, con l'obiettivo di farci guidare da esso e provare ad estrapolare alcune considerazioni preliminari; in tutti i restanti punti che compongono l'analisi, l'obiettivo è quello di rispondere a quelle domande che ci siamo posti nel momento in cui abbiamo scelto di svolgere questo progetto:

1a) Concentrandoci sul calcio italiano nel suo complesso (tenendo conto dunque dalla Serie A fino alla Serie D ed al campionato Primavera) e rappresentando tutti gli scambi di tutte le società a tutti i livelli, è possibile trarre già alcune conclusioni dal grafo risultante?

1b) Cosa succede se consideriamo invece soglie minime via via più alte riguardanti il valore di mercato dei giocatori trasferiti? C'è un regime di oligarchia che porta una cerchia molto ristretta di squadre ad accentrare verso di sé l'acquisto dei migliori giocatori 'depredando' le squadre di fascia bassa, o al contrario c'è una distribuzione più equilibrata degli acquisti di valore?

2) Esistono, all'interno del campionato di Serie A delle squadre "satellite" di altre dal punto di vista dello scambio di giocatori?

È possibile dunque rilevare delle sorte di "alleanze" tra squadre o gruppi di squadre che tendono a creare reti di scambi ricorrenti particolarmente fitte tra loro?

3) Quale percorso deve fare in Italia un calciatore per emergere dalle serie minori fino ad arrivare alla Serie A? Quanto è tortuosa la strada in termini di numero di squadre/leghe in cui passare prima di arrivare a destinazione? Quali sono le squadre che più fanno da 'ponte' verso il raggiungimento della massima serie?

4) Considerate le squadre del campionato italiano di Serie A, è possibile stabilire se una squadra è di alto livello, medio livello o basso livello, basandosi soltanto sulle caratteristiche dei calciatori acquistati come età, valore di mercato e campionato di provenienza? Quali skills portano un calciatore a trasferirsi verso una grande squadra? Quali verso una squadra media? Quali verso una squadra di basso livello?

DESCRIZIONE DATASET

Il dataset preso in considerazione conta 70.006 righe è disponibile al seguente link di GitHub:

github.com/d2ski/football-transfers-data 

Esso raccoglie tutti i trasferimenti delle top sette leghe europee citate nella premessa, e ad ognuna delle 70.006 righe corrisponde un trasferimento. Gli anni presi in considerazione vanno dal 2009 al 2021.

La fonte da cui il dataset attinge le informazioni è il sito Transfermarkt .

Transfermarkt è un sito web tedesco contenente informazioni calcistiche relative a calciatori, classifiche, risultati, squadre, ed appunto trasferimenti, che lo rende a pieno titolo un vero e proprio grande database calcistico. La fama di Transfermarkt è diffusa in Europa e nel mondo, e molti sono gli articoli di giornali sportivi che lo citano quando vengono snocciolati numeri e statistiche sul calcio.

Le colonne di cui è composto il dataset da noi utilizzato (le quali sono state rinominate per una maggiore comprensione), sono le seguenti;

- **season**: annata in cui avviene il trasferimento (esempio: 2015);
- **window**: la finestra di mercato in cui avviene il trasferimento. Può essere la finestra di calciomercato estivo (indicata con 's') o la finestra di calciomercato invernale (indicata con 'w');
- **team_id** : l'ID univoco di ciascuna squadra, attribuito da Transfermarkt;
- **team1**: il nome della squadra che cede il giocatore;
- **country_team1**: la nazionalità del team1;
- **league_team1**: la lega di appartenenza del team1;
- **dir**: la direzione del trasferimento. Se impostata su 'left' allora team1 è il club cedente e team2 l'acquirente, viceversa se è impostata su 'in'. Come specificato in seguito, è stato deciso di attribuire a tutte le righe con colonna dir impostata su 'in', il valore 'left', invertendo team1 e team2: in questo modo team1 sarà sempre il club cedente e team2 sempre il club acquirente;
- **player_id** : l'ID univoco di ciascun giocatore, attribuito da Transfermarkt;
- **player_name**: il nome del giocatore protagonista del trasferimento;
- **player_age**: l'età del giocatore al momento del trasferimento;
- **player_nation**: nazionalità del giocatore;
- **player_pos**: ruolo del giocatore;
- **team2**: il nome della squadra che acquista il giocatore;
- **team2_country**: la nazionalità del team2;
- **team2_league**: la lega di appartenenza del team2, la quale è stata aggiunta da noi;
- **transfer_value**: il valore monetario del trasferimento del giocatore, ovvero la somma pagata dalla squadra acquirente a quella cedente (EUR);

- **market_value:** il valore di mercato del giocatore al momento in cui il trasferimento avviene (EUR). Il valore di mercato è una stima effettuata da Transfermarkt sulla base di parametri quali età, ruolo, campionato di appartenenza, squadra di appartenenza, e altre statistiche;
- **is_free:** se 'True' identifica un calciatore acquistato "a parametro zero". Ciò significa che il trasferimento avviene quando il contratto con la squadra cedente è arrivato a scadenza. In questo caso chi acquista il giocatore lo acquista gratuitamente. Nel caso contrario, la colonna è impostata su 'False';
- **is_loan:** se 'True' significa che si tratta di un calciatore acquistato in prestito. Nel caso contrario, la colonna è impostata su 'False';
- **is_loan_end :** se 'True' identifica un calciatore di ritorno da un prestito. Si tratta del movimento inverso a quello che avviene per il prestito, di conseguenza ad ogni prestito corrisponderà un trasferimento inverso di fine prestito. Viceversa la colonna è impostata su 'False';
- **is_retired :** se 'True' significa che il calciatore si sta trasferendo dal suo club di appartenenza a nessun'altro club, in quanto è giunto il momento del suo ritiro. Viceversa la colonna è impostata su 'False';
- **transfer_id :** l'ID univoco di ciascun trasferimento, attribuito da Transfermarkt.

DATASET: IL PRE-PROCESSING

 Sulla stessa [pagina di GitHub](#) indicata in precedenza è disponibile il codice, utilizzato dai creatori del dataset, per effettuare lo scraping delle informazioni. Questo ci ha permesso di analizzare il codice e di adattarlo alle nostre esigenze.

Abbiamo quindi eseguito il codice relativo allo scraping ma adattandolo alle nostre esigenze. Nello specifico abbiamo:

- estratto anche gli anni 2007-2008, che si aggiungono agli anni dal 2009 al 2021 già presenti nel dataset. In questo modo è stato possibile allargare l'analisi agli ultimi 15 anni.
- estratto anche i trasferimenti relativi ai campionati italiani di Serie B, Serie C, Serie D, Campionato Primavera. Questo ci ha consentito di poter andare a fondo nel nostro focus sui trasferimenti delle squadre italiane.

Il dataset così ottenuto, nonostante assecondare le nostre esigenze in termini di anni e campionati da considerare, ci ha messo comunque di fronte a diverse sfide affrontate nella fase di pre-processing. Di seguito verranno descritti soltanto i passaggi che hanno consentito di risolvere le problematiche più evidenti:

- Nel dataset, come accennato precedentemente, i team coinvolti nel trasferimento sono team1 e team2. Se la colonna "dir" è impostata su "left", allora il giocatore passa da team1 a team2; se la colonna "dir" è impostata su "in" allora il giocatore passa da team2 a team1.

Abbiamo ritenuto che questo potesse creare delle ambiguità, di conseguenza abbiamo trasformato tutte le colonne "in" in "left", invertendo i valori di team1 e team2, come indicato in precedenza.

- Per la quasi totalità delle colonne i valori mancanti riscontrati sono stati pochissimi. Le uniche due colonne problematiche da questo punto di vista sono state il transfer value ed il market value, entrambe necessarie per svolgere la nostra analisi.

In particolare per il market value, non solo gli Na sono presenti in numero rilevante, ma abbiamo riscontrato che, per ogni trasferimento, il market value associato a quel trasferimento non è quello relativo al momento in cui il trasferimento avviene (come dovrebbe essere) ma quello relativo al valore di mercato attuale del giocatore (nel 2022).

Di conseguenza la colonna market value è stata completamente azzerata.

Per ricavare le informazioni relative ai market value ed ai transfer value mancanti ci siamo serviti di un secondo dataset, denominato da noi dataset_aggiuntivo.csv:

<https://github.com/emordonez/transfermarkt-transfers/tree/master/data> 

🔗 Questo dataset, al contrario del nostro, presenta meno righe e meno informazioni nelle colonne, ma le colonne market value e transfer value presentano la quasi assenza di valori nulli. Abbiamo dunque implementato un algoritmo che permettesse, per ogni riga del nostro dataset con valori mancanti nel transfer/market value, di individuarlo e prenderlo dal dataset aggiuntivo, facendo il match con il nome del giocatore, stagione, squadra di arrivo, squadra di partenza.

- La colonna relativa al campionato di appartenenza del team2 non esiste. Di conseguenza per ogni squadra della colonna team2, siamo andati a prelevare la lega di appartenenza di quella squadra quando la squadra è presente nella colonna team1, ricordando che per il team1 esiste la colonna relativa al campionato di appartenenza.
- Le squadre che negli anni presi in analisi sono fallite e state poi rifondate con ridefinizione, presentano diversi nomi ma riferiti in realtà alla medesima squadra. In questo caso abbiamo cercato, man mano che le incontravamo durante la nostra analisi, di fare in modo, il più possibile, di unificare queste denominazioni multiple in unica denominazione che indicasse quella squadra.

Legenda sulle leghe di appartenenza: sia per il team1 che per il team2, tramite il codice scritto da noi, abbiamo fatto in modo che:

- Le leghe squadre appartenenti alle top7 nazioni prese in considerazione escludendo l'Italia, cioè Germania, Inghilterra, Spagna, Francia, Portogallo, Olanda, fossero indicate rispettivamente:
 - a) con GER1, ENG1, SPA1, FRA1, POR1, NET1 se appartenenti al massimo livello calcistico di quella nazione.
 - b) con GER2, ENG2, SPA2, FRA2, POR2, NET2 se appartenenti alle serie calcistiche minori di quella nazione.
- le leghe delle squadre appartenenti a tutte le altre nazioni corrispondono alle prime 3 lettere del nome di quella nazione. Ad esempio: il Santos ha come nazione di appartenenza 'Brazil', in questo caso la sua lega di appartenenza sarà 'BRA'
- le leghe delle squadre appartenenti al sistema calcistico italiano fossero indicate rispettivamente con:
 - a) ITA1 per le squadre appartenenti alla Serie A
 - b) ITA2 per le squadre appartenenti alla Serie B
 - c) ITA3 per le squadre appartenenti alla Serie C
 - d) ITA4 per le squadre appartenenti alla Serie D
 - e) ITAJ per le squadre appartenenti al Campionato Primavera.

Il tutto viene eseguito in modo automatico basandosi sui dati già presenti nel dataset.

ANALISI ESPLORATIVA

Prima di partire con l'analisi vera e propria, si è deciso di effettuare una breve panoramica sui grandi numeri del calciomercato in Europa e sulla loro evoluzione in questi ultimi 15 anni.



Il primo grafico offre un riassunto dell'evoluzione del denaro speso nei trasferimenti da parte delle squadre dei top7 campionati europei.

Da segnalare l'anno 2009 in cui si è assistito ad un quasi raddoppio del denaro speso: è l'anno in cui entrano in azione gli sceicchi nel mondo del calcio, con l'Abu Dhabi United Group del principe emiratino Mansur bin Zayed Al Nahyan che rileva il Manchester City. Stesso discorso per il 2011, anno in cui il Paris Saint Germain viene rilevato dal qatariota Nasser Al Khelaifi.

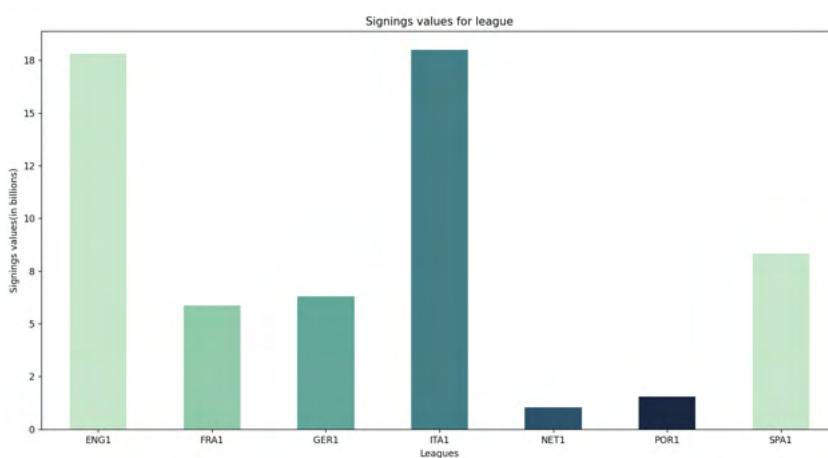
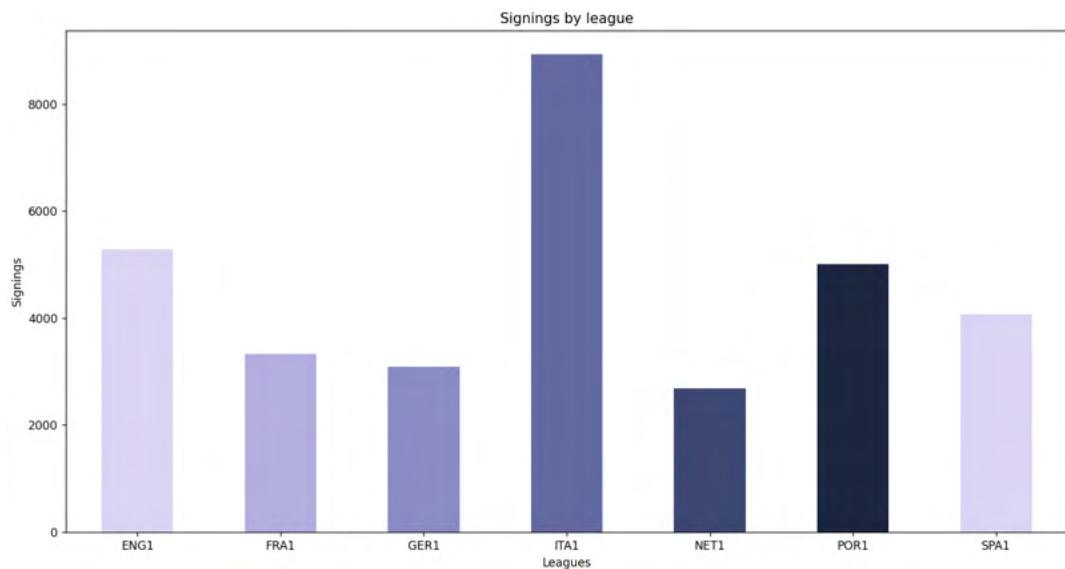
Dal 2014 fino al 2019 il trend è stato in forte e costante crescita. Si è passati da meno di 3 miliardi spesi nel 2014 ad addirittura i 7 miliardi spesi nel 2019.

Un segnale della sempre maggiore importanza delle finestre di calciomercato, capaci nel giro di poco più di due mesi di stravolgere le squadre, con i calciatori che hanno raggiunto valori sempre più alti.

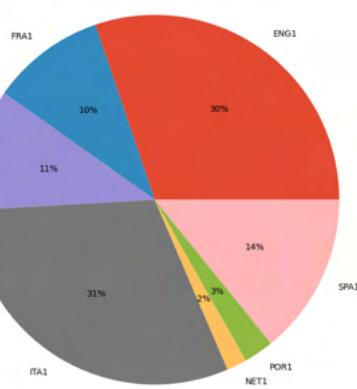
A sostegno di questo, la classifica dei trasferimenti più costosi della storia:

nelle prime 25 posizioni soltanto i passaggi di Neymar dal Santos al Barcellona (88mln) e di Gareth Bale dal Tottenham al Real Madrid (101 mln) sono stati effettuati prima del 2014. Per trovare un trasferimento antecedente al 2007 bisogna invece risalire fino alla 27esima posizione, con Zinedine Zidane che passò dall'italiana Juventus al Real Madrid per un ammontare di 77mln.

La bolla sembra essere scoppiata in concomitanza della diffusione della pandemia da covid19: nel 2020 e poi nel 2021 il giro di denaro è drasticamente sceso fino a tornare ad un livello di circa 3 miliardi di euro.



Signings amount by league



I due grafici a barre di sopra invece, rappresentano rispettivamente il numero di acquisti ed il denaro investito in calciomercato, negli ultimi 15 anni, da ciascuna delle top7 leghe prese in considerazione. Per i soldi investiti in calciomercato è rappresentato anche un grafico a torta che indica in percentuale quanto ogni lega contribuisca al totale dell'ammontare di denaro speso per i trasferimenti effettuati.

Da evidenziare, basandosi sui due diagrammi a barre, come l'Italia sia prima per distacco per numero di acquisti effettuati. Una distanza abissale che si azzerà invece se consideriamo non il numero di acquisti ma i soldi effettivamente investiti. In questo caso a dominare è la Premier League, con la Serie A che resta comunque di poco staccata in seconda posizione. Il confronto tra i due grafici denota come in Italia le squadre siano molto attive, ma allo stesso tempo i trasferimenti sono meno impattanti, dal punto di vista economico, rispetto a quelli della Premier League, che con meno transazioni riesce ad aggiudicarsi il primato come lega che spende maggiormente.

PRIMO PUNTO

Fin dal primo punto, come accennato nella premessa, abbiamo deciso di andare a fondo sul campionato italiano, considerando non solo la Serie A, ma tutti gli altri livelli professionistici (Serie B, Serie C), il maggior livello dilettantistico (Serie D), e il Campionato Primavera dei giovani.

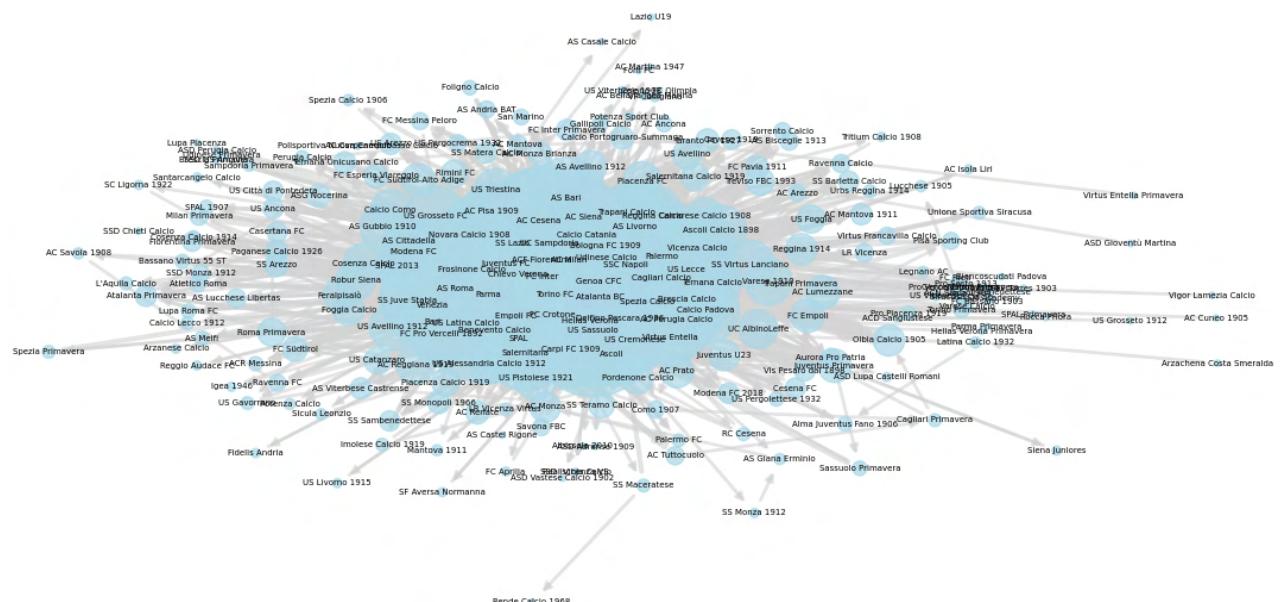
Abbiamo disegnato un grafo che rappresentasse tutti i trasferimenti avvenuti negli ultimi 15 anni nei 5 livelli appena elencati.

Nel disegnare il grafo abbiamo deciso inizialmente di non impostare alcuna soglia di filtraggio del dataset, in modo tale da vedere se questo potesse permetterci di fare delle prime considerazioni sul mercato in Italia.

Il grafo è stato realizzato utilizzando il **Kamada Kawai Layout**, che si basa sull'algoritmo di Kamada e Kawai.

Attraverso il suddetto algoritmo, i nodi che occupano una posizione centrale sono quei nodi che sono direttamente collegati, o meglio adiacenti, con la maggior parte degli altri nodi. I periferici, invece, sono quei nodi che presentano un'interconnessione minore, per cui la distanza media con tutti gli altri nodi, considerata la lunghezza dei percorsi minimi, è maggiore. In particolare, la distanza viene calcolata utilizzando la *shortest_path_length*. Di conseguenza i nodi centrali sono caratterizzati da un grado elevato rispetto ai nodi più periferici che presentano invece un minor numero di laterali adiacenti.

Tale layout è stato utilizzato anche nella costruzione di tutti gli altri grafi del progetto eccetto i grafi del terzo punto nel quale è stato utilizzato lo **Spring layout**.



Il grafo plottato in questo modo, potrebbe sembrare apparentemente insignificante. Soffermandoci sui nomi delle squadre tuttavia, ad occhio, abbiamo intuito potesse esserci una tendenza per la quale, le squadre posizioionate nella parte centrale del grafo sono anche quelle che solitamente sono presenti in Serie A, mentre via via, allargandoci per cerchi concentrici si passa sempre di più verso le squadre di serie minori(B, C, D, Primavera).

Si è deciso dunque di colorare le squadre in base alla loro appartenenza ad una delle cinque serie italiane prese in considerazione.

Occorre specificare che, ovviamente, nell'arco di 15 anni una stessa squadra, retrocedendo e venendo promossa, può aver militato in più serie diverse.

La nostra idea è stata quindi quella di aggiungere un ulteriore colonna al dataset che etichettasse ciascuna squadra italiana con la lega in cui più ha militato nei 15 anni presi in considerazione.

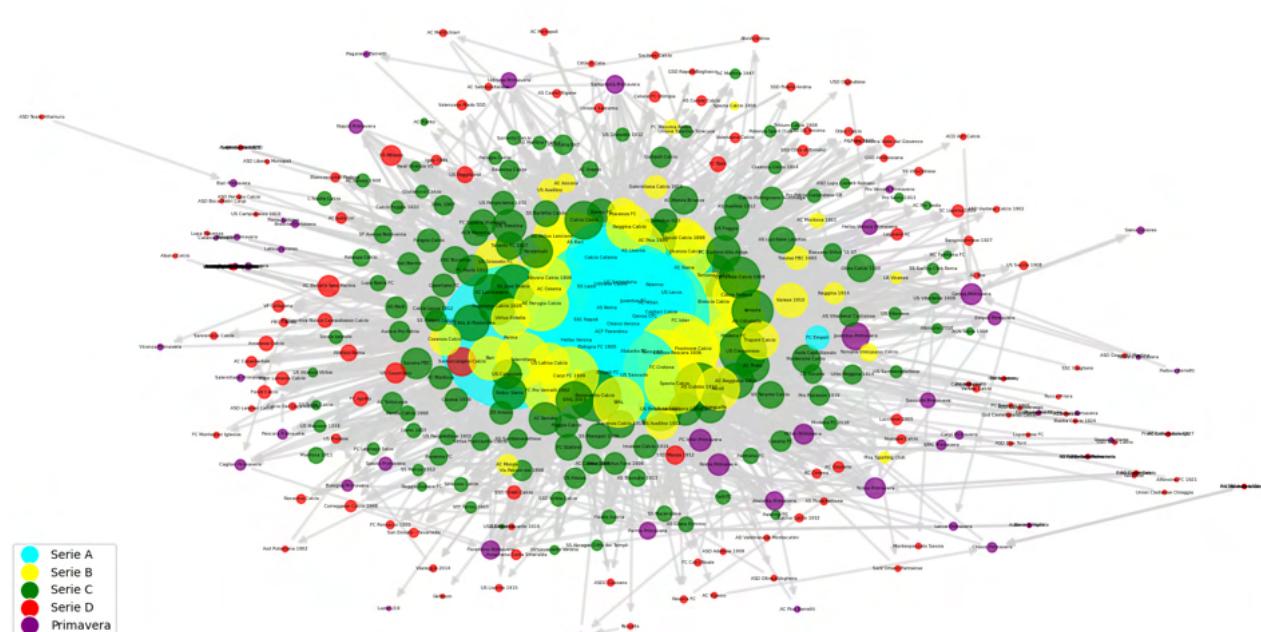
Per esempio, il Cagliari Calcio dal 2007 ad oggi, in 15 anni, ha partecipato a 14 campionati di Serie A e ad un solo campionato di Serie B, di conseguenza è stata etichettata come squadra di Serie A.

Il tutto è stato fatto andando a contare, per ogni squadra, quante volte nel corso degli anni è militata in ciascuna serie, andando poi a prendere la lega con il conteggio più alto.

Una volta colorati i nodi delle squadre sulla base della loro lega di appartenenza, il risultato è il seguente:

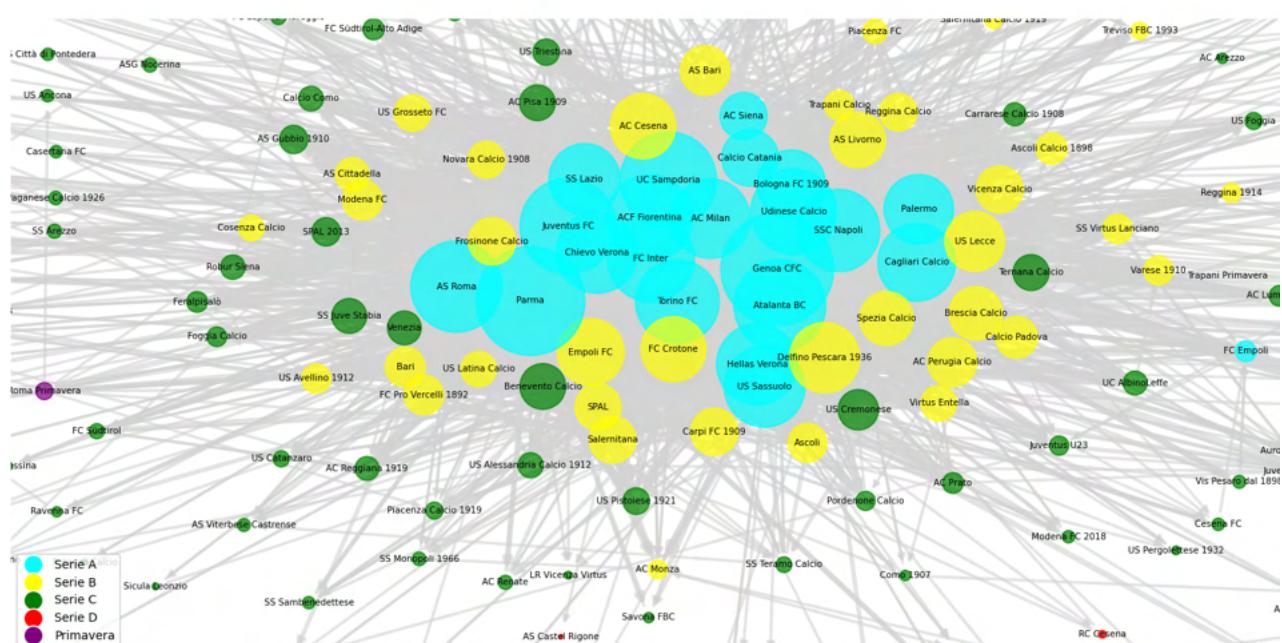
PRIMO GRAFO

Football transfers between italian clubs from 2007 to 2021



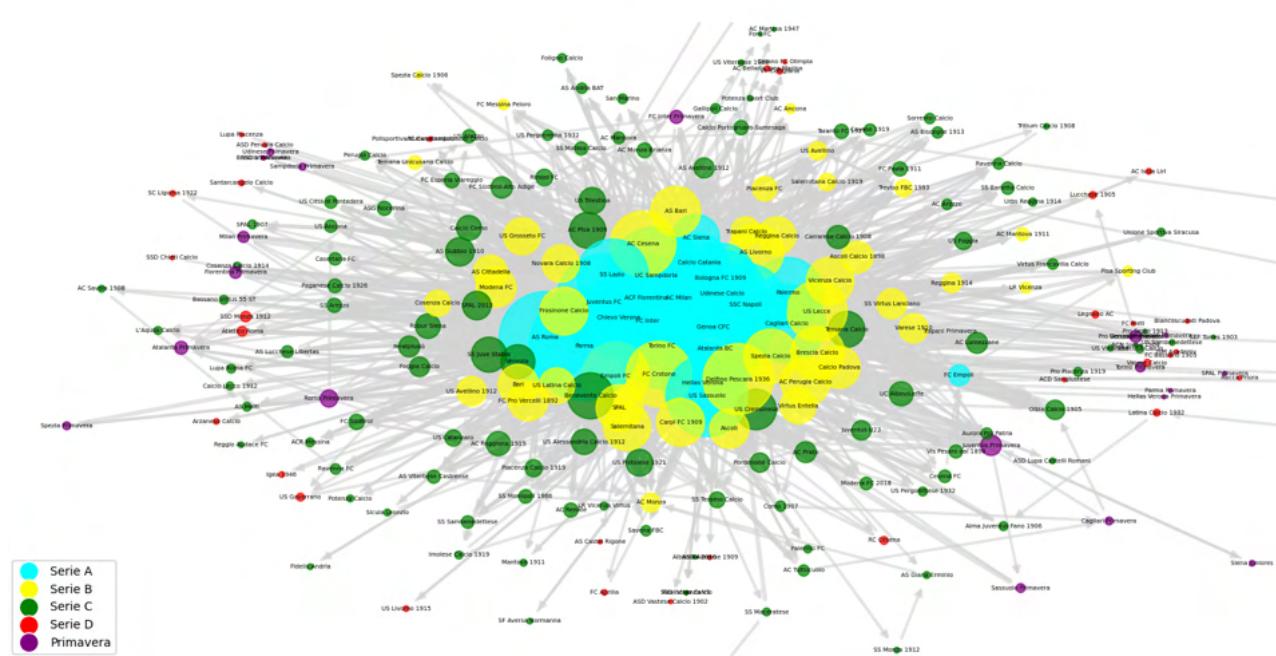
PRIMO GRAFO ZOOM

Football transfers between Italian clubs from 2007 to 2021

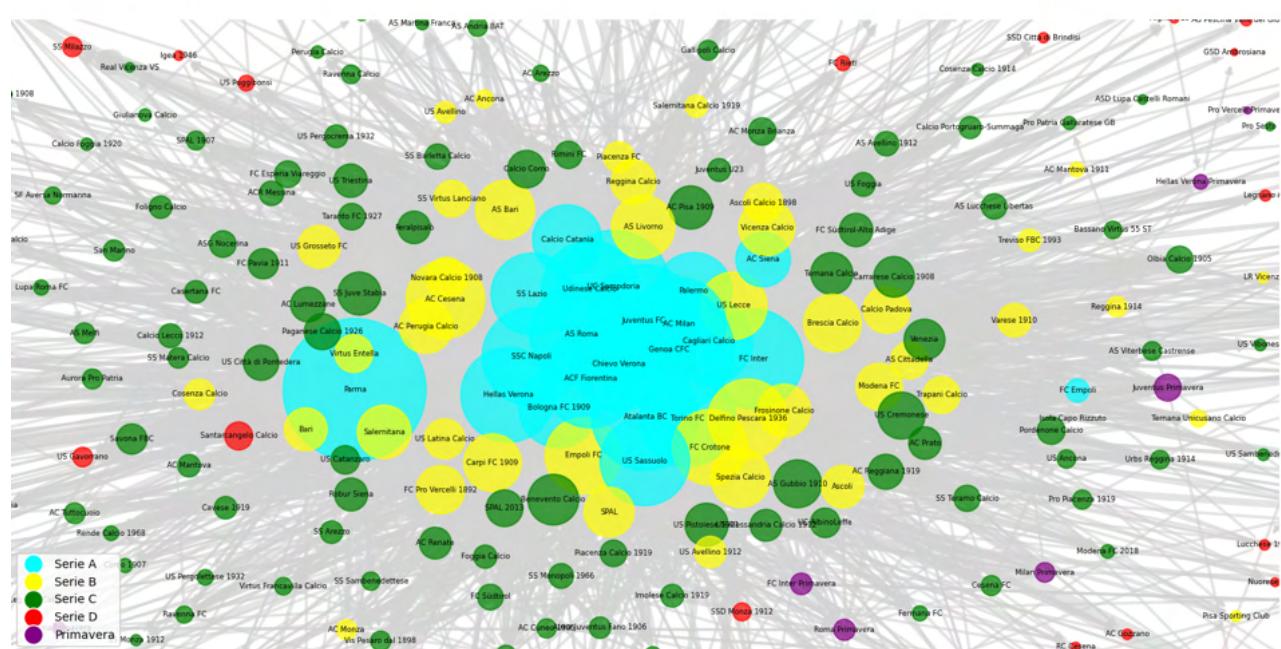


SECONDO GRAFO

Football transfers between Italian clubs from 2007 to 2021

**SECONDO GRAFO ZOOM**

Football transfers between Italian clubs from 2007 to 2021



Nel primo grafo non abbiamo impostato alcuna soglia, mentre nel secondo abbiamo impostato una soglia minima, seppur molto bassa, di market value, pari al primo quartile. Per ciascuno dei due grafi è stato effettuato anche lo zoom nella parte centrale, in modo che siano più visibili le squadre presenti.

Vediamo come sia possibile in entrambi i casi individuare dei cerchi concentrici quasi perfetti: nella parte più interna e centrale del grafo le squadre sono di Serie A, man mano che si va verso la periferia si scala di serie, fino ad arrivare alla parte più esterna dominata dalle squadre di Serie D e dalle giovanili del Campionato Primavera.

Ricordiamo che il Kamada Kaway layout fa sì che i nodi che occupano una posizione centrale nel grafo sono quei nodi direttamente collegati, o meglio adiacenti, con la maggior parte degli altri nodi.

Su questo presupposto, vediamo come i club di Serie A, seppur si stiano prendendo in considerazione i trasferimenti di tutti i club di tutte le serie, occupano in maniera interessante una posizione centralissima. Di conseguenza è possibile supporre che questi tendano a manovrare con le proprie operazioni i passaggi anche di gran parte dei giocatori che si muovono nelle serie inferiori. Man mano che invece ci si sposta verso la periferia del grafo (verso i cerchi concentrici più esterni e quindi verso le squadre prima di Serie B, poi di Serie C, e infine di Serie D e primavera) questo "potere" va scemando.

Tutto ciò potrebbe essere collegato alla strategia, da parte dei club italiani, di **vdersi accaparrare i giocatori a basso valore quando promettenti**, e quindi ad acquistarne il cartellino, per poi farli crescere attraverso continui prestiti degli stessi giocatori verso le serie minori, senza che questi, in molti casi, arrivino poi a giocare veramente con il club che sulla carta ne detiene il cartellino. Questa strategia permette ad un club di acquisire un giocatore e controllarne, attraverso i vari prestiti verso le squadre minori, tutti i trasferimenti, tenendo sempre in pugno il controllo della situazione.

Essendo che i prestiti contano come trasferimenti, si spiega come le squadre delle serie maggiori siano via via più centrate nel grafo rispetto alle serie minori; per esempio: se il Milan manda per 5 volte in prestito un giocatore verso squadre di Serie C, per il Milan conteranno 5 trasferimenti, mentre solo 1 per ciascuna delle 5 squadre di Serie C verso cui il giocatore sotto controllo viene mandato in prestito).

Di seguito riportiamo un esempio, che riguarda il calciatore Cristian Pasquato:

#11 Cristian Pasquato



1
1
1



Trento
Serie C - A
 Livello: Terzo livello
 In rosa da: 08/set/2021
 Contratto fino: 30/giu/2024

225 mila €

Aggiornato al: 20/giu/2022

TRASFERIMENTI						
Stagione	Data	Venditore	Acquirente	VdM	Costo	
21/22	08/set/2021	svincolato	Trento	250 mila €	-	>
21/22	01/lug/2021	Gubbio	svincolato	250 mila €	-	>
20/21	21/ago/2020	Campodarsego	Gubbio	125 mila €	gratuito	>
19/20	21/ott/2019	svincolato	Campodarsego	150 mila €	-	>
19/20	01/lug/2019	Legia Varsavia	svincolato	500 mila €	-	>
17/18	20/lug/2017	Juventus	Legia Varsavia	1,80 mln €	250 mila €	>
16/17	30/giu/2017	KS Samara	Juventus	1,80 mln €	Fine prestito	>
16/17	09/ago/2016	Juventus	KS Samara	1,00 mln €	Prestito	>
15/16	30/giu/2016	Pescara	Juventus	1,00 mln €	Fine prestito	>
15/16	01/feb/2016	Juventus	Pescara	1,30 mln €	Prestito	>
15/16	31/gen/2016	Livorno	Juventus	1,30 mln €	Fine prestito	>
15/16	31/ago/2015	Juventus	Livorno	1,40 mln €	Prestito	>
14/15	30/giu/2015	Pescara	Juventus	1,40 mln €	Fine prestito	>
14/15	01/set/2014	Juventus	Pescara	1,30 mln €	Prestito	>
14/15	01/lug/2014	Udinese	Juventus	1,60 mln €	1,50 mln €	>
13/14	30/giu/2014	Padova	Udinese	1,60 mln €	Fine prestito	>

Nel caso di Pasquato, vediamo come la Juventus lo acquisti dall'Udinese nella stagione 2014/2015 ad un costo di soli 1,6mln di euro. Da quel momento in poi vi sono addirittura quattro prestiti consecutivi verso Pescara, Livorno, ancora Pescara, e KS Samara, il tutto senza che il giocatore abbia mai effettivamente giocato con la Juventus, proprietaria del suo cartellino.

Proprio su questo aspetto, la FIFA ha deciso di porre rimedio, limitando l'utilizzo della formula dei prestiti nei trasferimenti. Sull'articolo riportato da Sky Sport Italia il 20 gennaio 2022 è possibile leggere:

"Dall'1 luglio 2022 la Fifa introdurrà un nuovo regolamento sui prestiti con l'obiettivo di sviluppare i giovani giocatori, **promuovere l'equilibrio competitivo e prevenire l'accaparramento dei calciatori**. Ogni squadra avrà un numero massimo di giocatori da poter prestare o da acquistare in prestito (8 da luglio 2022 che diventeranno 7 nel 2023 e 6 dal 2024 in poi)".

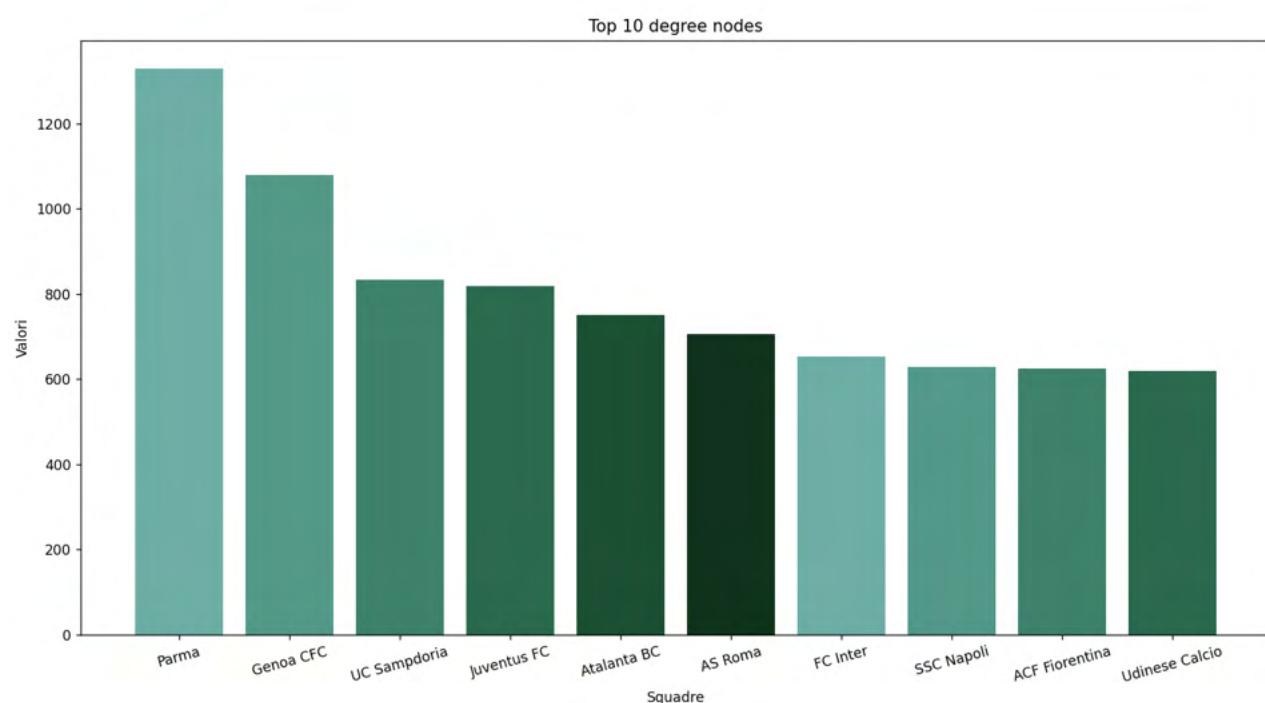
Relativamente al primo grafo abbiamo inoltre calcolato la degree dei nodi (e quindi delle squadre) e dunque la dimensione di ogni nodo all'interno del grafo è proporzionale alla propria degree.

Il grado di un nodo è il numero di connessioni dello stesso con gli altri nodi presenti nel network ed è una misura di centralità del nodo. Le misure di centralità sono delle misure che servono a cogliere e determinare l'importanza di un nodo, e il grado è una delle più semplici misure di centralità.

In tale contesto occorre fare una precisazione, che varrà per il resto del lavoro: nel presente progetto abbiamo lavorato, eccetto nel punto 4, con multigrafi orientati: nei multigrafi orientati 2 nodi possono essere collegati più volte tra loro stessi, sia in una direzione che nella direzione opposta, ed un nodo può essere collegato anche con se stesso. Inoltre, abbiamo fatto in modo che se, ad esempio un nodo si collega ad un altro nodo con 30 archi in una specifica direzione allora quest'ultimi verranno accorpati da un unico arco di peso 30 (es. se abbiamo 30 collegamenti dal nodo A al nodo B e 20 collegamenti dal nodo B al nodo A, risulterà un collegamento dal nodo A al nodo B di peso 30 ed un collegamento dal nodo B al nodo A di peso 20).

Di conseguenza, nel nostro multiDiGraph il grado di un nodo sarà la somma dei pesi del numero di archi in entrata ed in uscita da quel nodo.

Di seguito il diagramma a barre con la rappresentazione delle dieci squadre con degree più alta:

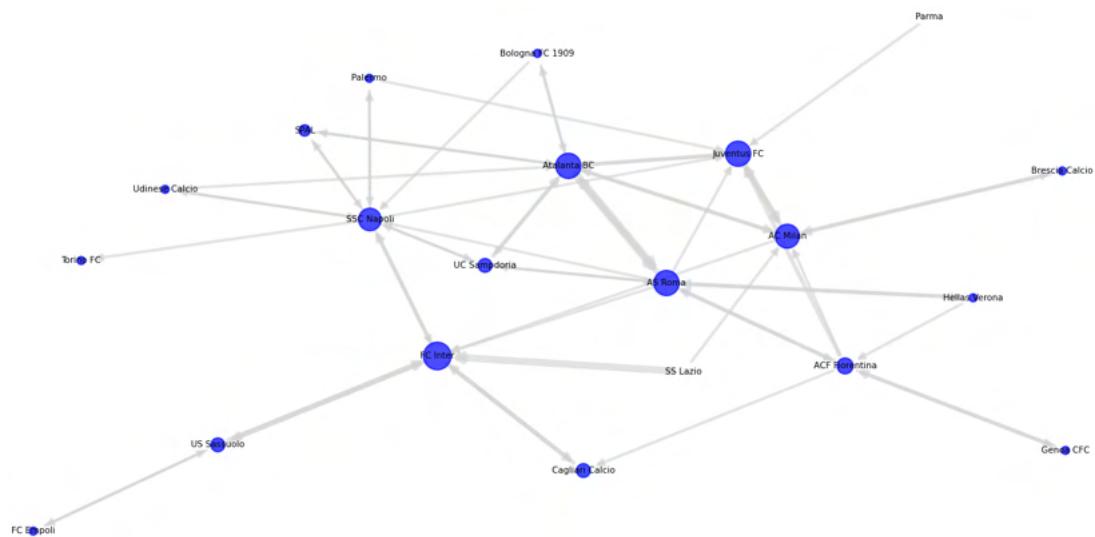


Vediamo come i 10 club con la più alta degree, appartengano tutti alla Serie A, confermando quanto detto sopra e quanto già visibile nel grafo.

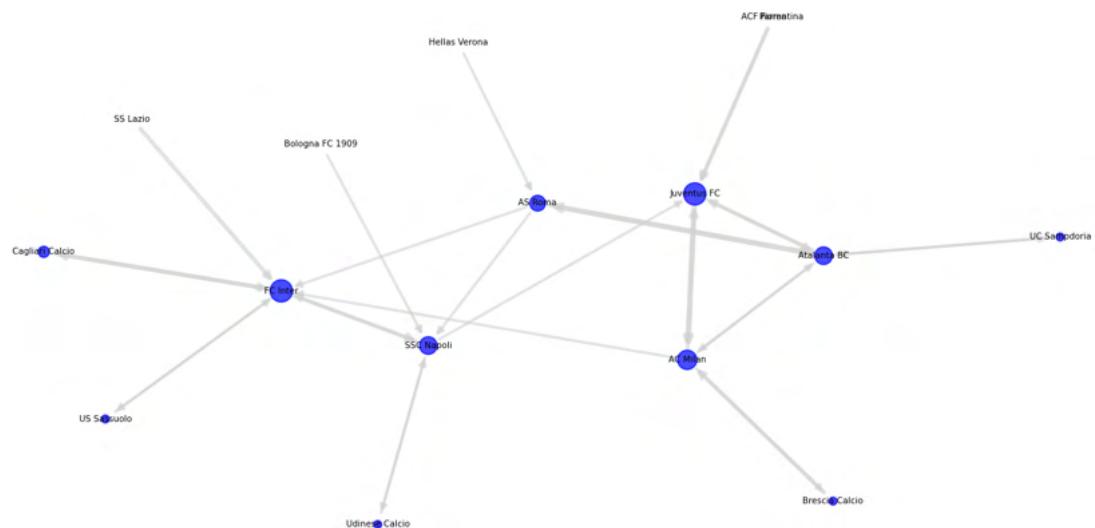
Arrivati a questo punto è interessante, con il medesimo grafo visto finora, andare ad innalzare via via la soglia del market value. Questo permette di verificare se esiste una sorta di oligarchia che porta una cerchia molto ristretta di squadre ad accentrare verso di se l'acquisto dei migliori giocatori 'depredando' le squadre di fascia bassa, o al contrario c'è una distribuzione più equilibrata negli acquisti dei giocatori con un valore rilevante.

Abbiamo impostato tre diverse soglie: 20mln, 25mln, 30mln.

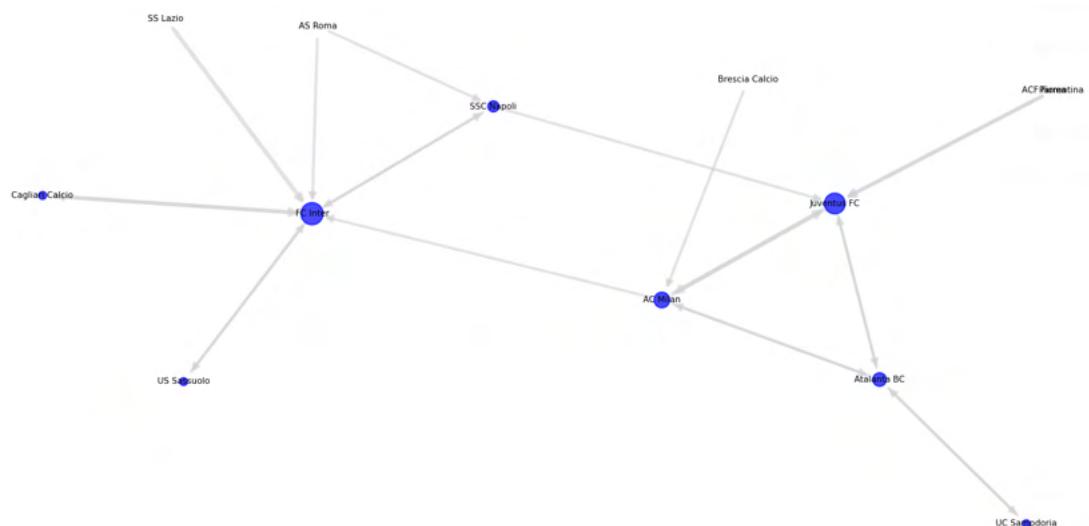
Football player transfer from 2007 to 2021 with filter market value (20 mln)



Football player transfer from 2007 to 2021 with filter market value (25 mln)



Football player transfer from 2007 to 2021 with filter market value (30 mln)



Una rappresentazione simile tramite i grafi è particolarmente utile in quanto consente, cambiando di volta in volta la soglia, di vedere come si evolve la situazione, dando all'utente un' informazione immediata dal punto di vista visivo.

In questo caso specifico possiamo vedere come, aumentando le soglie, vengano scremate pian piano le squadre acquirenti (verso cui arrivano gli archi in entrata) ed allo stesso tempo la direzione degli archi permette di capire già a primo impatto visivo quelle che sono, per ogni squadra acquirente, le squadre che più contribuiscono alla vendita di giocatori di pregio verso di essa, ovvero quelle squadre che ne costituiscono i serbatoi dai quali attingono maggiormente i giocatori di valore.

Le squadre acquirenti spiccano in quanto visibili con il nodo più grande. Se nei grafi precedenti infatti la grandezza del nodo si basava sul suo degree, stavolta, proprio per evidenziare le squadre che acquistano i calciatori di valore, abbiamo optato per basare le grandezze del nodo sull'in-degree.

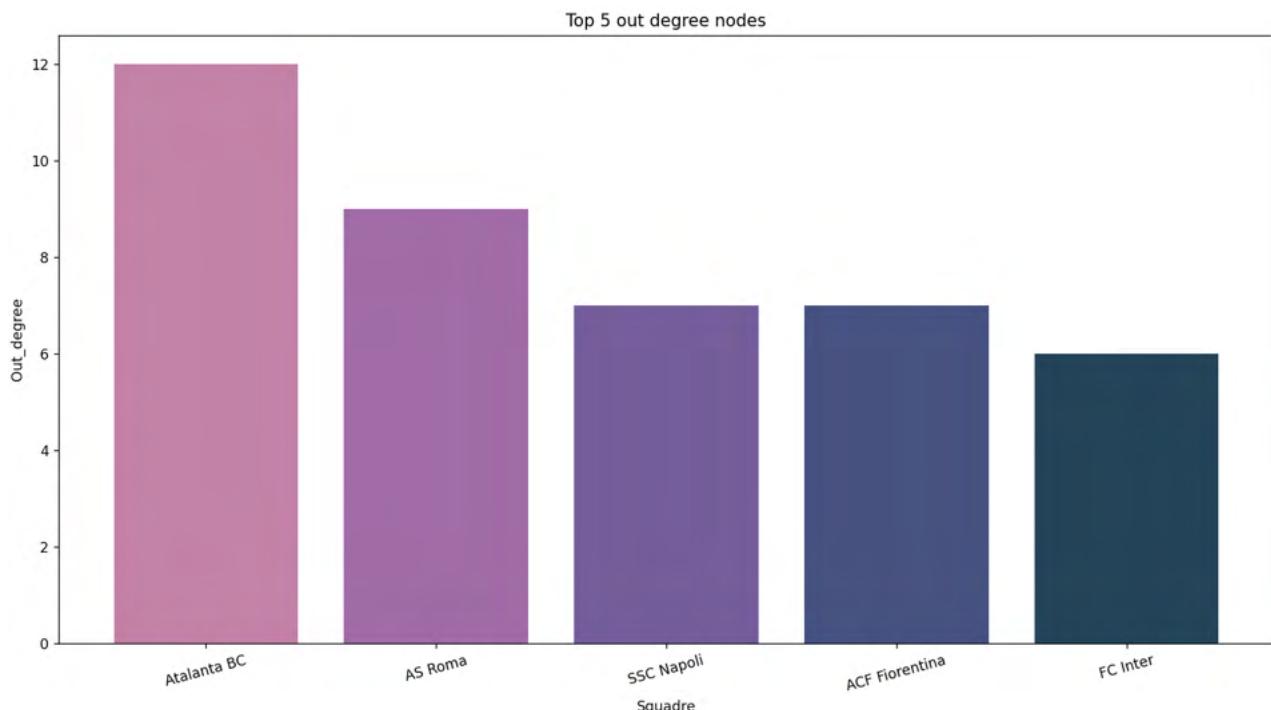
L'in-degree considera esclusivamente gli archi in entrata verso il nodo (e quindi gli acquisti) mentre non tiene conto degli archi in uscita (le vendite).

Con la prima soglia impostata a 20mln spiccano già solamente 6 squadre: Juventus, Milan, Inter, Napoli, Roma, Atalanta.

Per ciascuna di essa è possibile individuare i maggiori 'fornitori':

- Per la Juventus: Parma, Milan, Roma, Palermo, Atalanta, Udinese
- Per l'Inter: Sassuolo, Cagliari, Napoli, Lazio, Roma
- Per il Milan: Lazio, Fiorentina, Brescia, Juventus, Atalanta
- Per l'Atalanta: Sampdoria, Roma, Bologna, Spal
- Per il Napoli: Inter, Spal, Palermo, Bologna, Sampdoria, Udinese
- Per la Roma: Atalanta, Verona, Fiorentina, Sampdoria

Da notare come l'Atalanta sia presente come "venditrice" in 3 delle 5 squadre "acquirenti" (esclusa ovviamente l'Atalanta stessa). Questo il grafico delle top5 squadre con l'out degree più elevato:

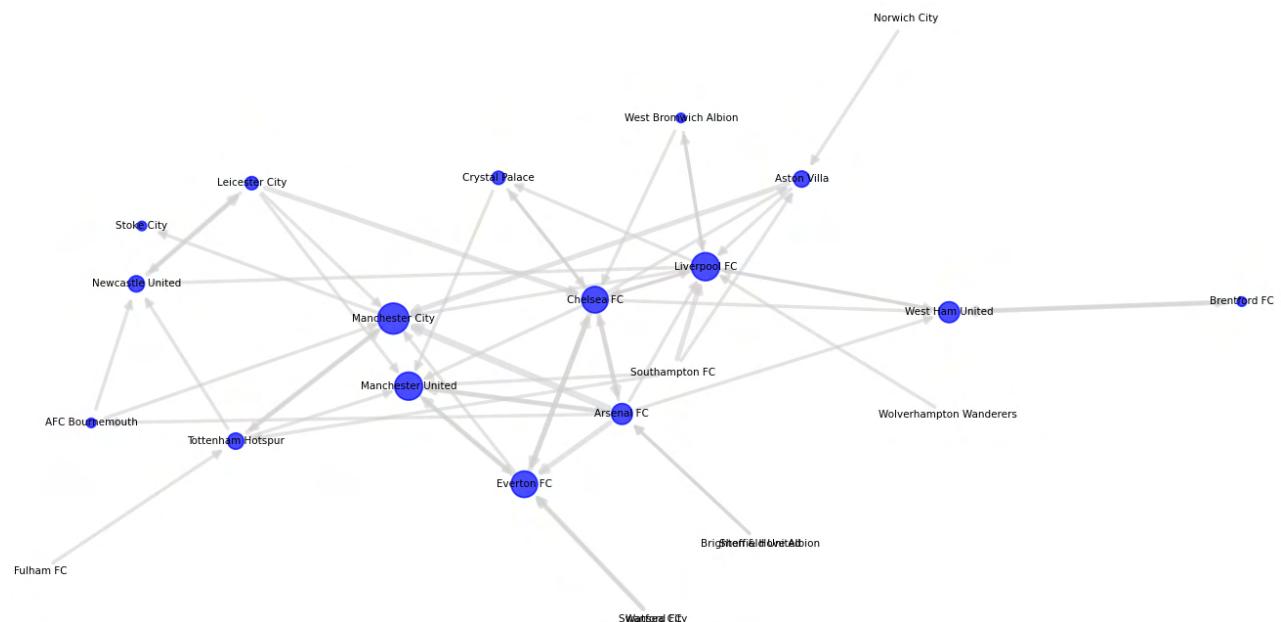


L'out degree, al contrario dell'in degree, misura il numero di connessioni in uscita in un nodo. Le squadre con l'out degree più elevata sono quelle che più hanno compiuto trasferimenti di valore verso le altre squadre del medesimo campionato, e l'Atalanta si trova in prima posizione.

Aumentando la soglia a 25 mln e poi 30mln resistono le 3 storiche 'big' del nostro campionato: Juventus, Inter, Milan, accompagnate dal Napoli.

Se ripetiamo lo stesso grafo con la soglia a 20mln anche per i campionati inglese, spagnolo, e tedesco notiamo che:

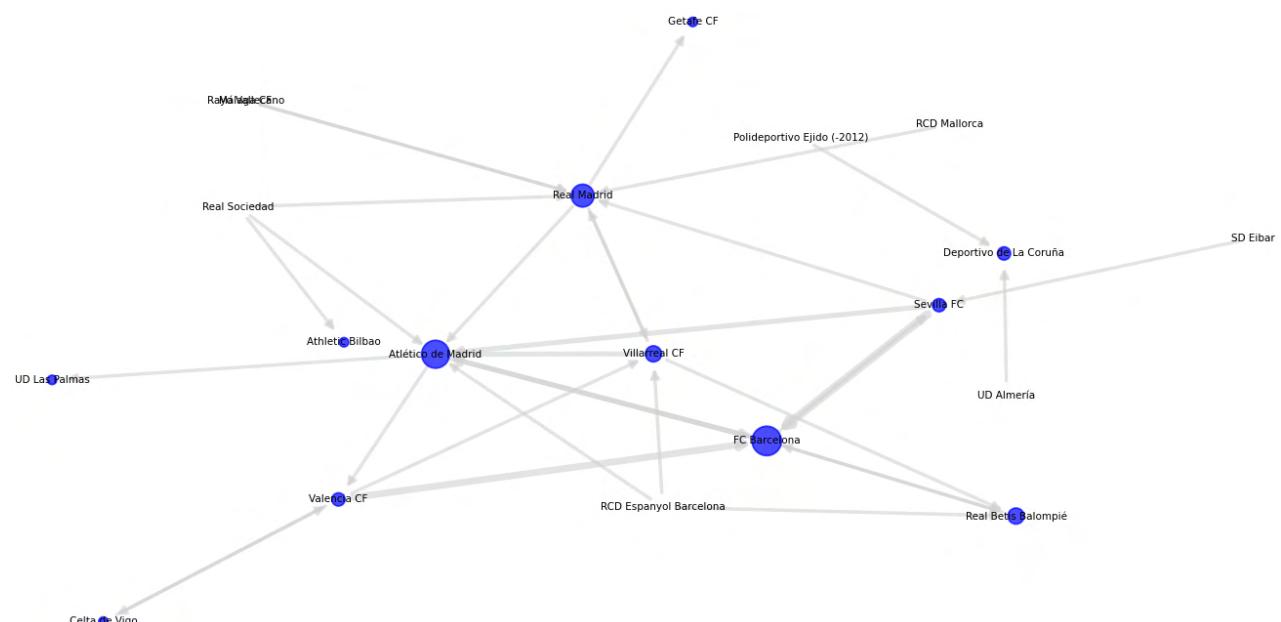
Football player transfer from 2007 to 2021 with filter market value (ENG)



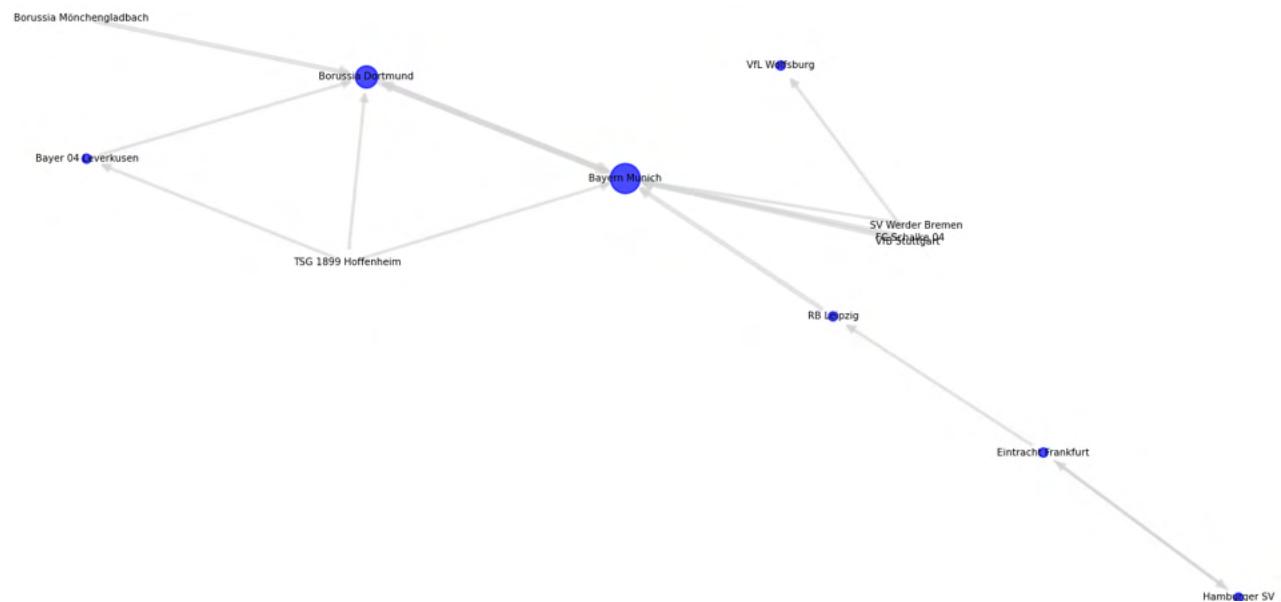
In Inghilterra la situazione è più o meno simile all'Italia. Le squadre "predatrici" sono sempre sei, ovvero Manchester United, Manchester City, Everton, Chelsea, Arsenal, Liverpool.

La situazione è simile nonostante in Inghilterra, come visto nella panoramica, si tendano a fare acquisti più costosi. La spiegazione potrebbe essere legata al fatto che, più che saccheggiare le squadre del proprio campionato, i club inglesi tendano ad acquistare i calciatori di valore dai campionati esteri.

Football player transfer from 2007 to 2021 with filter market value (ES)



Football player transfer from 2007 to 2021 with filter market value (DE)



In Spagna e Germania c'è invece una cerchia molto più ristretta di squadre che tendono a depredare le proprie connazionali. In particolare, il campionato tedesco è praticamente un duopolio tra Bayern Monaco e Borussia Dortmund.

SECONDO PUNTO

Il secondo obiettivo del lavoro è di verificare se sia possibile individuare all'interno della Serie A delle alleanze tra squadre, ovvero coppie o gruppi di squadre che tendono a creare nel tempo reti di scambi ricorrenti particolarmente fitte tra loro.

Per fare questo cercheremo di individuare e di separare nel nostro grafo le varie comunità di squadre, rientrando quindi nell'ambito della cosiddetta **community detection**.

La community detection (detta anche graph partition) è uno degli aspetti fondamentali nell'ambito della network analysis, in cui l'obiettivo è trovare gruppi di nodi che sono, in un certo senso, più simili tra di loro rispetto agli altri nodi.

"Il concetto di community detection è emerso nella network science come metodo per individuare specifici gruppi (comunità) all'interno di sistemi complessi, attraverso la loro rappresentazione mediante un grafo."

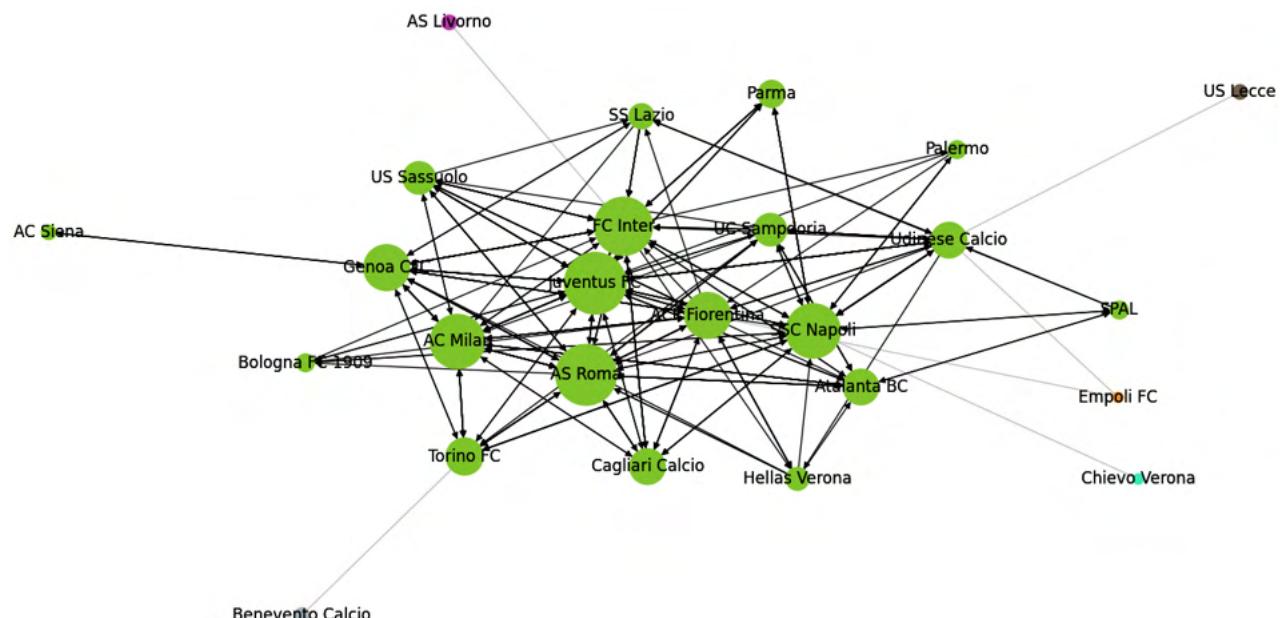
I metodi di rilevamento della comunità trovano sottoretti con un numero statisticamente significativo di collegamenti tra nodi nello stesso gruppo rispetto a nodi in gruppi diversi (Girvan e Newman, 2002)."



Nel nostro lavoro abbiamo operato con due algoritmi di community detection: *il Girvan-Newman algorithm*, che fa parte degli algoritmi di hierarchical community detection, ed il *Greedy modularity algorithm*, che fa parte degli algoritmi della modularity-based community detection.

L'algoritmo Girvan-Newman è stato il primo ad essere utilizzato, ma non ha fornito risultati soddisfacenti:

Communities with Girvan Newman



Di conseguenza abbiamo optato per il Greedy modularità algorithm, di cui riporteremo gli output ottenuti. Di seguito una descrizione di entrambi gli algoritmi.

Girvan-Newman algorithm

L'algoritmo Girvan-Newman implementato per la community detection si basa sull'iterativa eliminazione degli archi attraversati dalla maggiore quantità di flusso informativo che passa per sentieri minimali* che collegano tutte le possibili coppie del grafo. Di conseguenza vengono rimossi quegli edge che maggiormente fanno parte degli shortest path (sentieri minimali) che collegano le coppie di nodi del grafo. Rimuovendo gli edge del grafo uno per uno, questo si scomponete in parti più piccole, chiamate appunto comunità. L'algoritmo è stato sviluppato da Michelle Girvan e Mark Newman, dai quali eredita il nome. La ratio dell'algoritmo è di trovare quali archi nel grafo sono presenti maggiormente nei sentieri minimali che collegano i nodi, attraverso la misurazione dell'edge betweenness centrality. Questo perché ci si aspetta che i nodi che uniscono le diverse comunità abbiano un'elevata edge betweenness, e di conseguenza una volta eliminati gli edge con la betweenness più alta le comunità saranno molto più facili da individuare.

L'algoritmo Girvan-Newman può essere suddiviso in 4 step principali:

1. Per ogni edge del grafo, calcola l'edge betweenness centrality;
2. Rimuove l'edge con la betweenness centrality più elevata;
3. Calcola la betweenness centrality di ogni edge rimanente;
4. Ripete gli step dal 2 al 4 fino a quando non rimane nessun edge.

*Per sentiero minima si intende il sentiero più breve possibile che collega una coppia di nodi.

Greedy modularity algorithm



È un algoritmo di community detection che si basa sulla modularità. Quest'ultima rappresenta il criterio guida dell'algoritmo per la scelta e suddivisione delle comunità, che cerca appunto di ottenere il valore ottimale di modularità, massimizzandola.

La modularità è una metrica utilizzata per misurare la qualità di un particolare partizionamento del grafo o rete, e quindi è una misura utilizzata per misurare la bontà del partizionamento dei nodi. Essa quantifica la forza della comunità di un grafo comparando la frazione di edges all'interno della comunità con la frazione di edges che si avrebbe se venissero effettuate connessioni causali tra i nodi.

La ratio sottostante, spostandosi sul paragone umano, è che una comunità, se è tale, dovrebbe avere più collegamenti tra le persone che la compongono (nodi membri della comunità) rispetto ad un raduno casuale di persone.

La modularità può assumere valori compresi tra 0 e 1; un valore della modularità vicino allo 0 significa che la frazione di nodi all'interno della comunità non è migliore rispetto all'ipotesi casuale, mentre un valore della modularità pari ad 1 significa che la struttura della comunità ha la massima forza possibile. Dunque, l'obiettivo nella community detection che utilizza questa metrica è di trovare dei gruppi di nodi che abbiano la modularità più alta possibile.

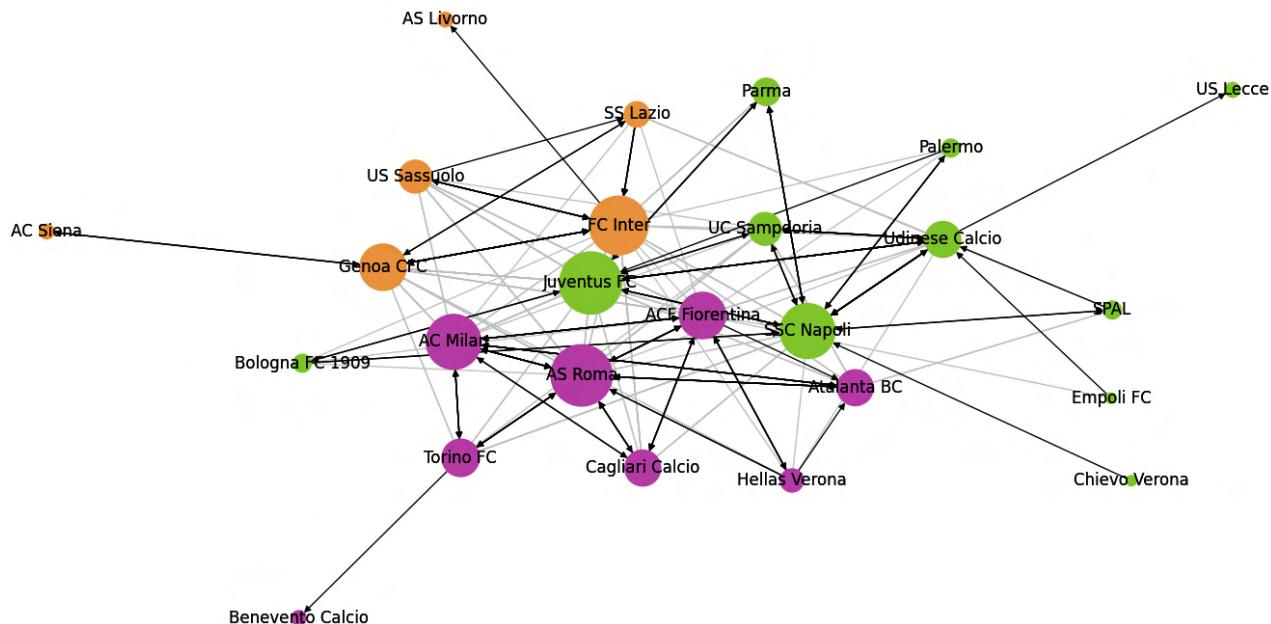


La strategia dell'algoritmo prevede che si parta con ogni nodo che inizialmente fa parte della sua comunità, e iterativamente l'algoritmo unisce le coppie di comunità che combinandosi portano al livello più alto di modularità, il tutto fino a quando non si raggiunge il limite massimo di modularità. È possibile, a discrezione dell'utente, impostare due condizioni di arresto alternative al raggiungimento del limite massimo di modularità: una condizione prevede di specificare il limite inferiore di comunità che si deve raggiungere per interrompere il processo, eventualmente anche ancor prima che il processo stesso raggiunga la modularità massima. La condizione opposta invece prevede di impostare un limite superiore che faccia sì che il processo si interrompa fino a quando rimangono al massimo n comunità, anche se è stato eventualmente già raggiunto il massimo della modularità.



Per analizzare le "alleanze" tra le squadre abbiamo preso in considerazione il decennio appena trascorso, dal 2010 al 2020, impostando una soglia minima di market value pari a 10 mln:

Communities with Modularity



Si può vedere come l'algoritmo di community detection individui tre comunità distinte di squadre. La comunità verde occupa la parte del centro-destra del grafico, la comunità arancione la parte centro-sinistra del grafo, la comunità viola la parte in basso.

Dal grafo emergono alcuni aspetti interessanti:

1. Le tre storiche big ed eterne rivali del calcio italiano, ovvero Juventus, Inter, e Milan, appartengono ciascuna a una comunità diversa.
2. Nel grafo abbiamo deciso di mettere in evidenza le squadre con maggiori acquisti, di conseguenza la grandezza dei nodi è proporzionale all'in-degree. Possiamo notare che in ciascuna delle 3 comunità, sono proprio Juventus, Inter e Milan a farla da padrone. Il Milan in realtà si contende l'egemonia con la Roma, anch'essa tra le squadre più blasonate del panorama italiano.
3. Nel calcio italiano professionistico esistono 5 derby, ovvero cinque casi di squadre che appartengono alla stessa città:
 - il derby di Milano tra Inter e Milan
 - il derby di Torino tra Juventus e Torino
 - il derby di Roma tra Lazio e Roma
 - il derby di Genova tra Sampdoria e Genoa
 - il derby di Verona tra Hellas Verona e Chievo Verona

Per ognuno di questi derby le squadre si trovano, in tutti e cinque i casi, in fazioni (comunità) opposte.

TERZO PUNTO

Il terzo obiettivo è quello di rispondere al seguente quesito: quale percorso deve fare, in Italia, un calciatore per emergere dalle serie minori (Primavera, Serie D, Serie C) ed arrivare alla Serie A? Nel rispondere a questo quesito andremo a verificare quanto sia tortuosa la strada in termini di squadre/leghe da attraversare prima di arrivare a destinazione, e quali sono le squadre che più fanno da ‘ponte’ nel raggiungimento del traguardo.

Per poter ottenere i grafi, abbiamo apportato alcune modifiche al dataset in modo che i dati presi in considerazione per l’analisi rispettassero 4 condizioni che di seguito vengono elencate. Le 4 condizioni sono descritte prendendo come esempio l’analisi del percorso dei giocatori che partono dalla Serie D, ma la stessa analisi verrà ripetuta con Serie C e Campionato Primavera:

- **PRIMA CONDIZIONE:** vengono presi in considerazione soltanto i giocatori il cui primo trasferimento, in ordine temporale nel dataset, ha come lega di partenza la Serie D (o C e Primavera a seconda di quale delle tre analisi stiamo considerando).
- **SECONDA CONDIZIONE:** dopo la scrematura della PRIMA CONDIZIONE vengono presi in considerazione soltanto quei giocatori che almeno una volta nella loro carriera arrivano in Serie A.
- **TERZA CONDIZIONE:** una volta che il calciatore arriva in Serie A, tutti i trasferimenti che in ordine temporale sono successivi al suo approdo in Serie A non vengono più presi in considerazione.
- **QUARTA CONDIZIONE:** l’approdo di un calciatore in Serie A viene considerato ‘valido’ soltanto se il calciatore permane in Serie A per almeno una stagione e mezzo, senza che questi venga immediatamente riceduto verso le serie minori. Di conseguenza, se entro la stagione e mezzo il calciatore ritorna in una serie inferiore, la TERZA CONDIZIONE viene annullata e si proseguirà nel tenere in considerazione i trasferimenti successivi.

Una volta modificato il dataset in base alle esigenze di analisi, per rispondere alle nostre domande, abbiamo deciso di dividere l’analisi in due parti:

- una parte in cui operiamo a livello macro, nella quale per ogni trasferimento prendiamo in considerazione la lega di partenza e la lega di arrivo. In questo caso il grafo sarà composto da 6 nodi ovvero il nodo Primavera, il nodo Serie D, il nodo Serie C, il nodo Serie B, il nodo Serie A, ed il nodo Estere che raggruppa i trasferimenti verso leghe estere. Ogni nodo accoppi al suo interno tutte le squadre che appartengono alla specifica lega rappresentata da quel nodo.
- una parte in cui operiamo a livello micro nella quale per ogni trasferimento prendiamo in considerazione la specifica squadra di partenza e di arrivo.

Per la parte macro, abbiamo calcolato per ogni analisi anche il numero di giocatori che nei 15 anni presi in considerazione sono riusciti effettivamente a raggiungere la Serie A.

A livello di grafo invece abbiamo impostato la grandezza dei nodi proporzionale alla loro degree e con la stessa logica dei punti precedenti. Gli archi in uscita rappresentano i calciatori che trasferendosi lasciano quella lega, gli archi in entrata i calciatori che trasferendosi approdano in quella lega. Anche in questo caso, tutti gli archi in entrata in ciascun nodo sono raggruppati in unico arco in entrata, il cui peso sarà dato dalla somma degli archi in entrata, analogamente per gli archi in uscita, dunque lo spessore di ogni arco è proporzionale al suo peso.

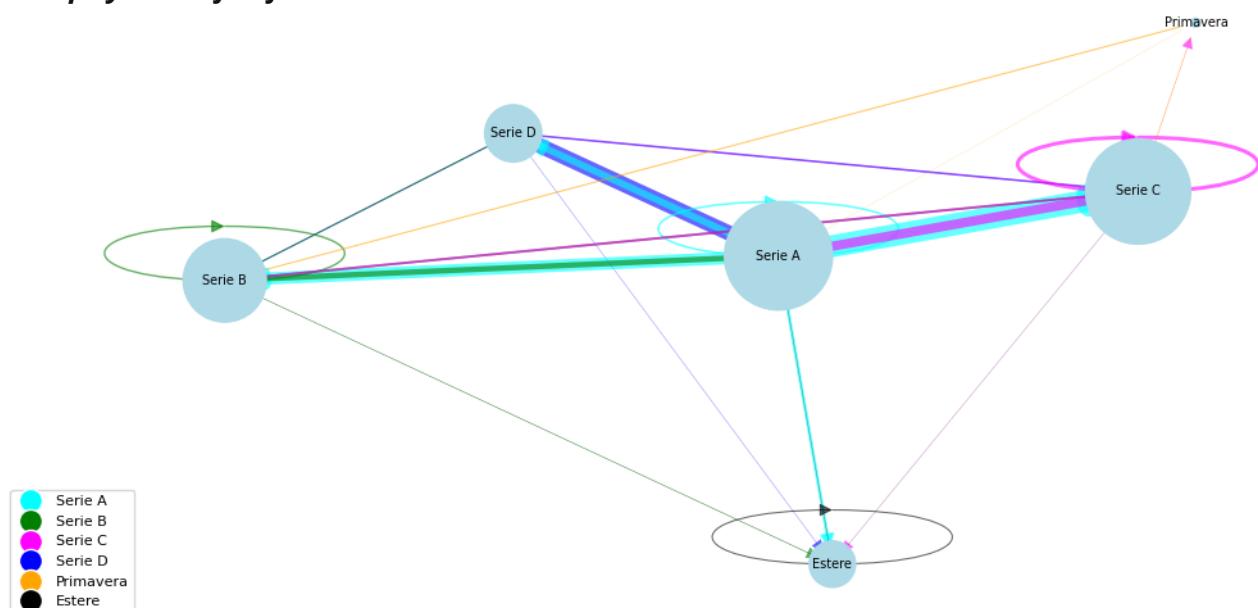
Per facilitare la comprensione dei grafi inoltre, abbiamo distinto i colori degli archi in uscita per ogni nodo del grafo, con questa legenda:

- Archi in uscita da Serie A: celeste
- Archi in uscita da Serie B: verde
- Archi in uscita da Serie C: fucsia
- Archi in uscita da Serie D: blu
- Archi in uscita da Primavera: arancione
- Archi in uscita da Estere: nero

Di seguito i grafi ottenuti:

Da Serie D a Serie A

Path player transfers from serie D to serie A



Nel caso della Serie D, soltanto 68 giocatori in 15 anni riescono a raggiungere la Serie A.

Ricordando che gli archi in uscita dalla Serie D sono di colore blu scuro, dal grafo vediamo, a sorpresa, che l'arco più grosso in uscita dalla Serie D è proprio verso la Serie A.

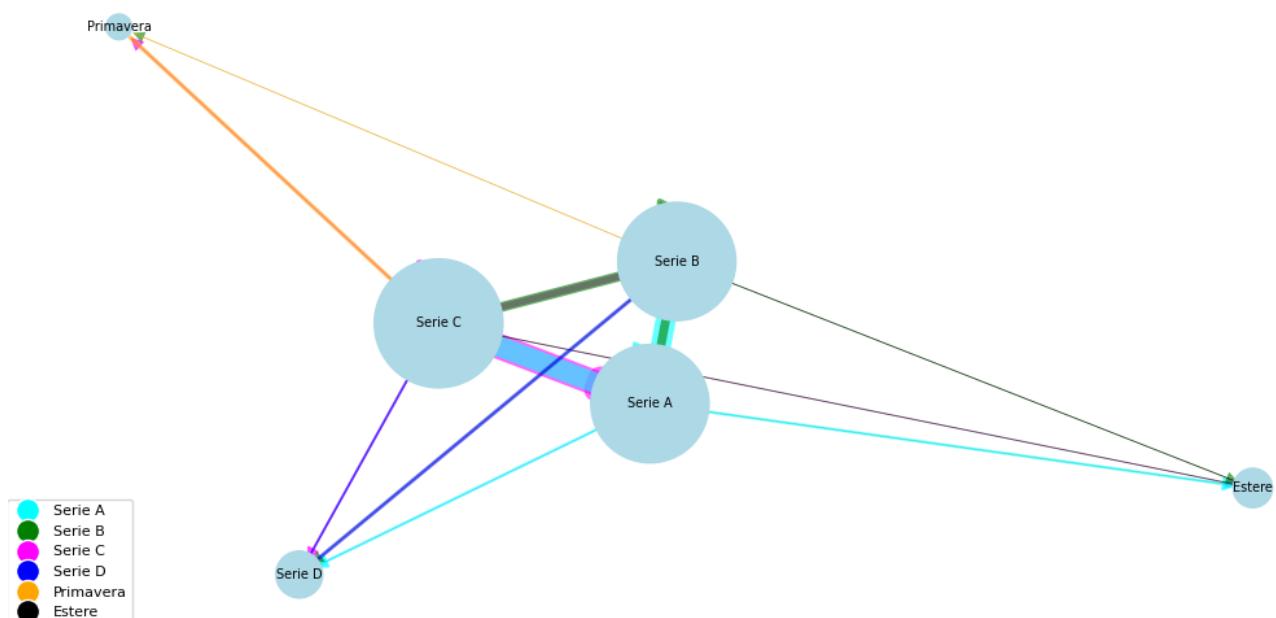
Allo stesso tempo però è possibile notare come la stessa Serie A abbia archi in uscita (celesti) di spessore notevole sia verso la Serie B, sia verso la Serie C, sia verso la stessa Serie D.

Questi due aspetti sottolineano come in realtà per un giocatore di Serie D non sia poi così difficile raggiungere la Serie A, ma come poi in realtà siano tantissimi i casi in cui, una volta raggiunta, la si lasci nel giro di una stagione e mezzo (ricordiamo infatti che per la QUARTA CONDIZIONE l'approdo in Serie A è valido solo se il giocatore permane per almeno una stagione e mezzo; se la QUARTA CONDIZIONE non è soddisfatta non ci si arresta, ma si continuano ad analizzare le tappe successive all'abbandono "prematuro" della Serie A).

Il fatto che tanti giocatori raggiungano direttamente la Serie A dalla Serie D, ma che altrettanto tanti vengono poi rispediti verso le serie minori entro 18 mesi, dovrebbe avvalorare la tesi già espressa al punto uno, per la quale le squadre di Serie A tendono molto spesso ad accaparrarsi i giocatori senza che poi questi effettivamente giochino in Serie A.

Per quanto riguarda gli altri archi blu uscita dalla Serie D, risultano tutti molto più sottili rispetto a quelli che entrano in Serie A. Nonostante gli archi in uscita dalla Serie D verso la Serie B e la Serie C siano come detto sottili, i nodi serie B e Serie C risultano comunque abbastanza grandi: la grandezza dei nodi Serie B e Serie C si spiega con il fatto che molti dei calciatori presi in considerazione in realtà transitano dalla Serie B e dalla Serie C arrivandoci però non dalla Serie D, ma dopo essere stati rimandati indietro dalla Serie A.

Path player transfers from serie C to serie A



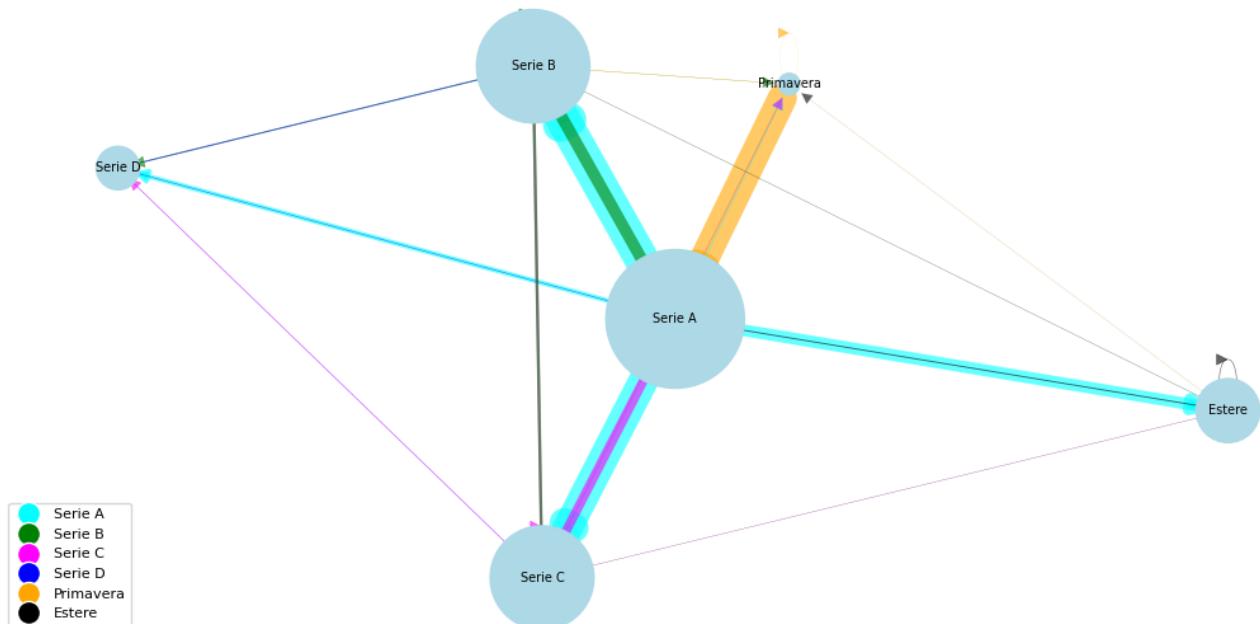
Nel caso della Serie C abbiamo numeri simili alla Serie D: i giocatori che in 15 anni sono riusciti effettivamente a raggiungere stabilmente la Serie A sono 78.

Anche in questo caso, come per la Serie D, tra i nodi in uscita della Serie C (fucsia) il più grosso è quello che va diretto alla Serie A. Allo stesso tempo, così come per la Serie D, sono tanti i giocatori che dalla Serie A fanno marcia indietro (archi celesti) entro 18 mesi, in particolare tornando in Serie C o andando in Serie B, ma alcuni addirittura effettuano un downgrade finendo in Serie D.

La situazione analizzata con la Serie D e con la Serie C è quindi molto simile e valgono dunque le medesime considerazioni fatte poco sopra.

Da evidenziare, rispetto al grafo della Serie D, come stavolta entri in gioco, sia in entrata che in uscita, il Campionato Primavera, prima poco collegato con il resto dei nodi. È un segnale di come tra la D e la Primavera non ci siano praticamente legami, questi ultimi invece esistono tra C e Primavera.

Path player transfers from Primavera to serie A



Nel caso della Primavera i numeri si alzano notevolmente: i giocatori che in 15 anni partono dalla Primavera e raggiungono stabilmente la Serie A sono 416, circa sei volte in più rispetto a Serie D e Serie C.

Ricordando che gli archi in uscita dalla Primavera sono arancio, vediamo che la quasi totalità dei calciatori anche in questo caso passa dalla Primavera alla Serie A. Tanti di essi vengono poi rispediti verso le Serie minori, soprattutto verso la Serie B e verso la Serie C.

Considerando gli archi in uscita dalla Serie B (verdi) e dalla Serie C (fucsia) vediamo che una buona parte riescono poi a tornare nuovamente in Serie A.

Per un giovane del campionato primavera dunque, sembrerebbe che il percorso più comune sia un passaggio immediato dalla Primavera alla Serie A, un ritorno verso la Serie B e la Serie C per proseguire la gavetta, e infine il salto definitivo nuovamente verso la Serie A.

Una volta analizzati i grafi, l'obiettivo finale è quello di calcolare rispettivamente:

- nel primo grafo, la lunghezza media di tutti sentieri che ogni calciatore percorre nel suo cammino dal nodo Serie D al nodo Serie A . Avremo quindi un sentiero per ogni singolo calciatore che parte dalla Serie D e riesce a raggiungere la Serie A.
- nel secondo grafo, la lunghezza media di tutti i sentieri che ogni calciatore percorre nel suo cammino dal nodo Serie C al nodo Serie A. Avremo quindi un sentiero per ogni singolo calciatore che parte dalla Serie C e riesce a raggiungere la Serie A.
- nel terzo grafo la lunghezza media di tutti i sentieri che ogni calciatore percorre nel suo cammino dal nodo Primavera al nodo Serie A. Avremo quindi un sentiero per ogni singolo calciatore che parte dalla Primavera e riesce a raggiungere la Serie A.

Calcolando questa lunghezza media dei sentieri, andiamo ad ottenere per ogni grafo il numero di passaggi che in media, ciascun giocatore, impiega per arrivare in Serie A.

Un sentiero è una successione di nodi tale che una coppia consecutiva è rappresentata da nodi vicini (dove per nodi vicini si intendono due nodi collegati da un arco) e la lunghezza del sentiero è data dal numero di archi che compone quel sentiero.

Abbiamo deciso di calcolare la lunghezza media di tutti i sentieri (che sono uno per ogni calciatore) nel seguente modo:

per ogni coppia di nodi, facciamo la somma del peso dell'arco in entrata e dell'arco di uscita del primo nodo della coppia, ed in questo modo otteniamo la somma di tutti i passaggi che tutti i calciatori passati dalla Serie D (o Serie C, o Primavera) alla Serie A hanno eseguito. Dopodiché dividiamo il risultato ottenuto per il numero di calciatori.

In questo modo otteniamo appunto quanti passaggi, in media, ogni giocatore effettua per arrivare in Serie A

Esempio esplicativo:

- supponiamo di avere soltanto i nodi Serie C, Serie B, Serie A
- I calciatori che partono dalla C ed arrivano alla A sono in tutto 100
- 50 passano direttamente da C ad A
- 25 passano da C a B e poi da B ad A
- 25 passano da C a B, tornano in C da B, e da C passano in A

Le coppie di nodi saranno CA, CB, BA

- Per la coppia CA abbiamo un peso di 75 (dato dal peso di 75(50+25) dell'arco in uscita da C verso A sommato al peso di 0 degli archi in entrata in C da A)
- Per la coppia CB abbiamo un peso di 75 (dato dal peso di 50(25+25) dell'arco in uscita da C a B sommato al peso di 25 dell'arco in entrata in C da B)
- Per la coppia BA abbiamo un peso di 25 (dato dal peso di 25 dell'arco in uscita da B ad A sommato al peso di 0 dell'arco in entrata in B da A)

Avremo quindi un totale di $75+75+25= 175$ passaggi, che diviso per i 100 giocatori fa una media di 1,75 passaggi a giocatore prima di raggiungere la Serie A.

Abbiamo ottenuto come valori in termini di trasferimenti medi:

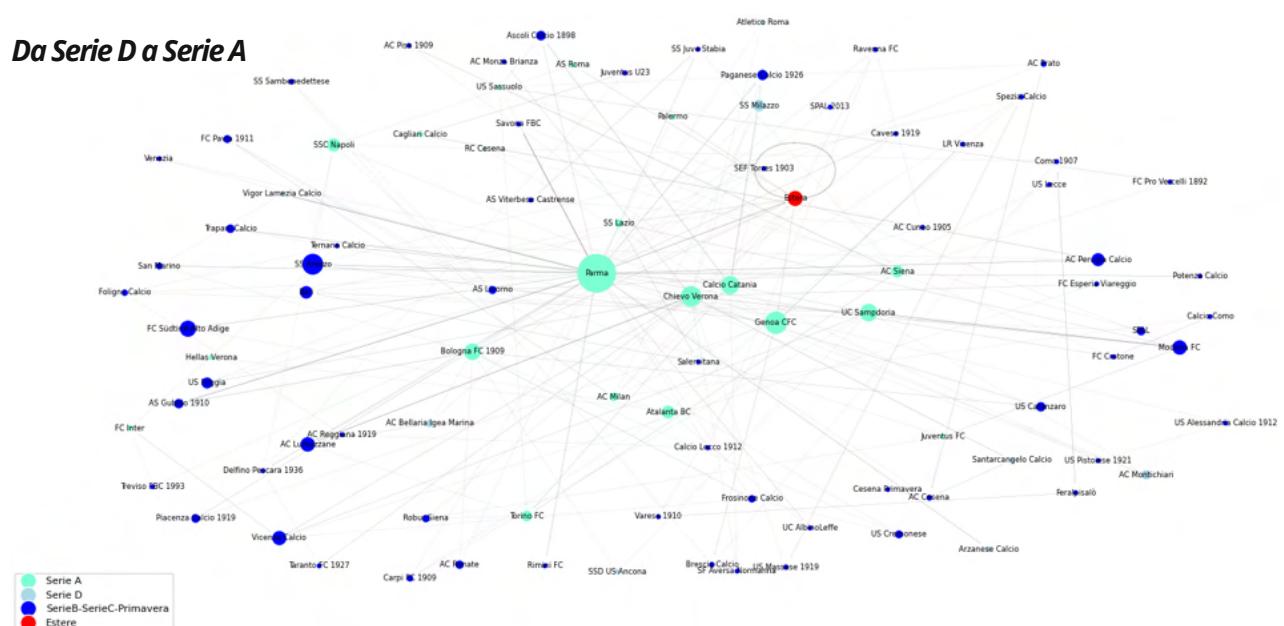
- 5,42 per la Primavera;
- 6,50 per la Serie D;
- 7,71 per la Serie C

Nella parte a **livello micro** prendiamo invece in considerazione le specifiche squadre che i calciatori attraversano nel percorso per arrivare alla Serie A.

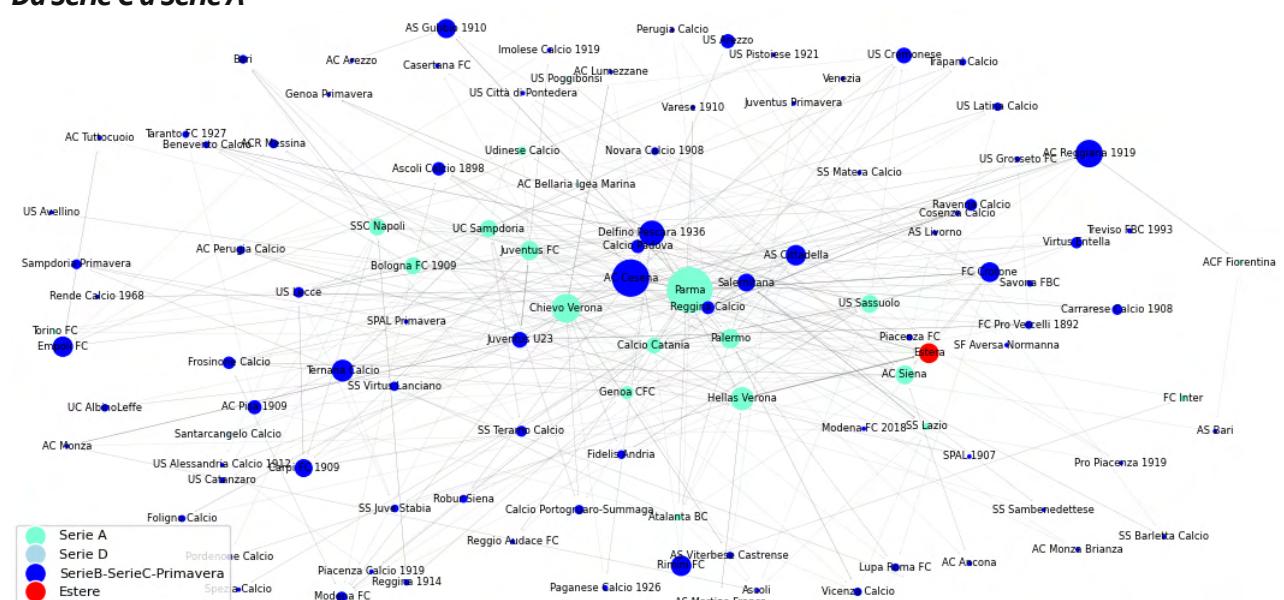
Lo scopo di questa parte è di capire quali sono le squadre che maggiormente fanno da 'ponte' più rapido per i calciatori che partono dalle serie minori fino ad arrivare al massimo livello del calcio italiano, ovvero quelle squadre attraverso le quali un calciatore transitandoci, tende in media ad abbreviare il suo percorso di raggiungimento della Serie A.

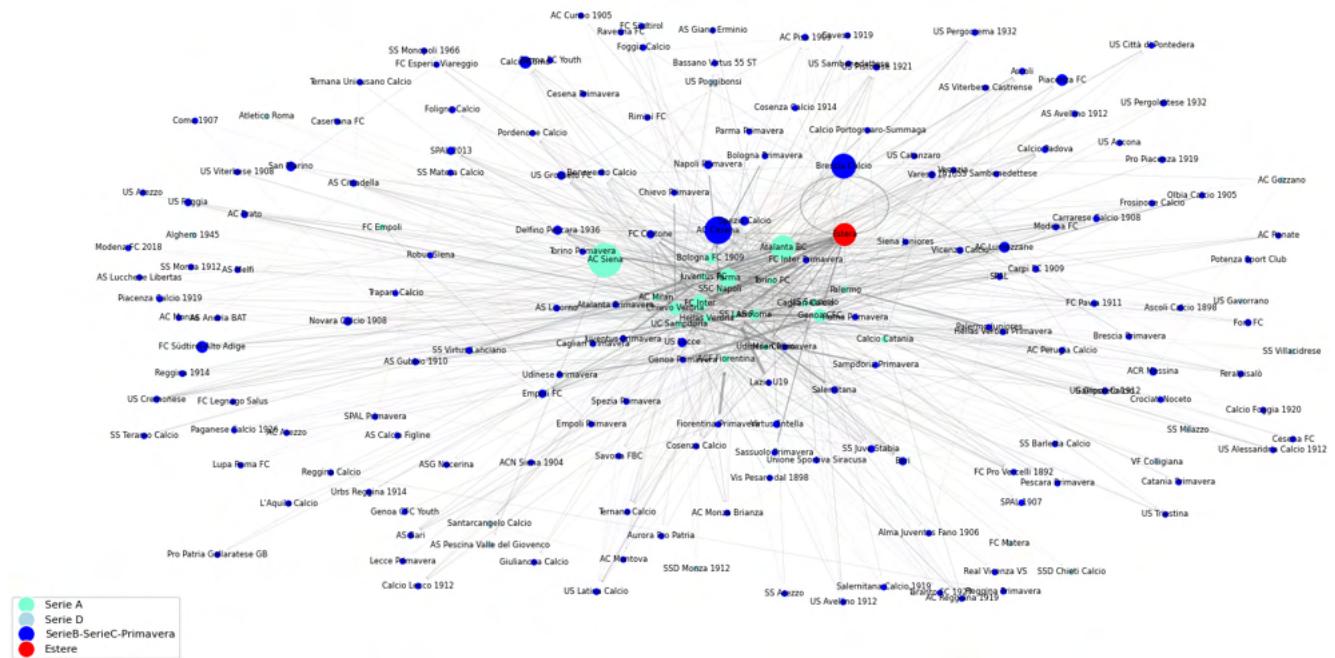
Anche in questo caso avremo tre grafi: uno per i calciatori che partono da squadre di Serie D, uno per i calciatori che partono da squadre di Serie C, ed uno per i calciatori che partono da squadre Primavera:

Da Serie D a Serie A

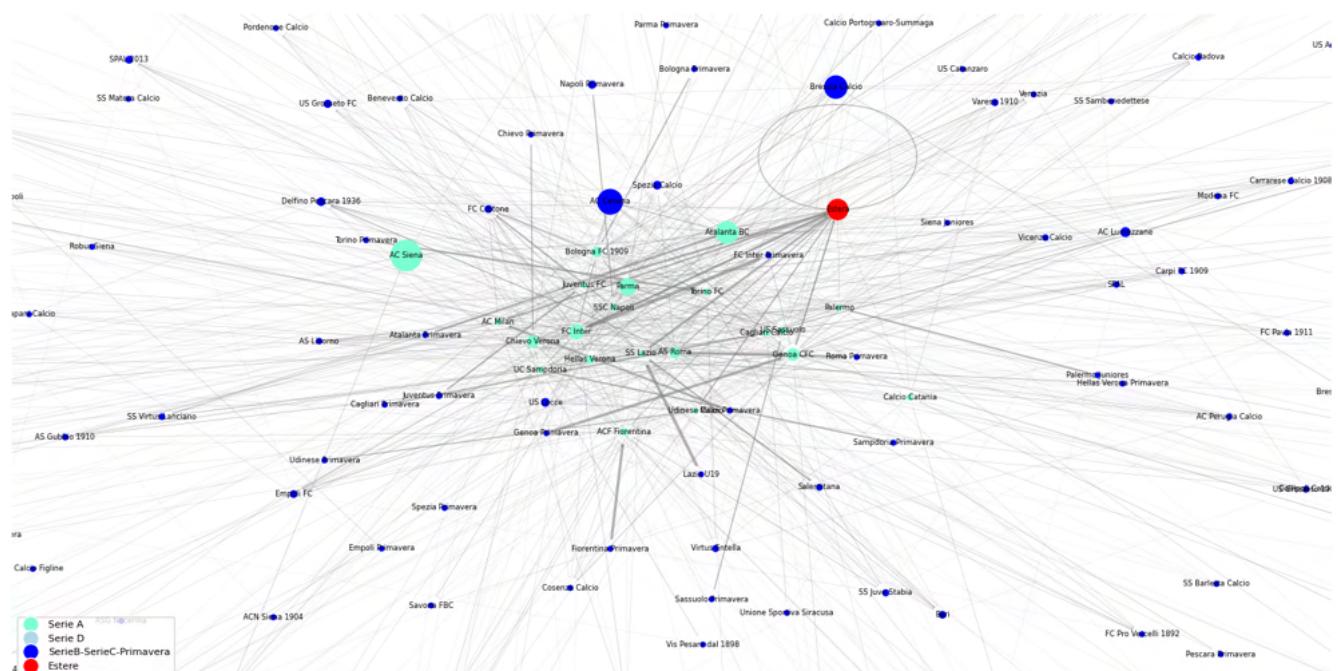


Da Serie C a Serie A



Da Primavera a Serie A

Per questo grafo, data l'eccessiva numerosità di nodi e l'eccessiva densità nella parte centrale, abbiamo effettuato anche uno zoom nella parte più interna del grafo:



In questo caso ciò su cui ci siamo concentrati è la betweenness centrality supponendo che, come verrà spiegato meglio in seguito, le squadre i cui nodi risultino con la più alta betweenness centrality siano quelle squadre che permettono ai calciatori di accorciare il loro percorso dalle Serie minori verso la Serie A.

Nei grafi dunque, la grandezza dei nodi stavolta è direttamente proporzionale alla loro betweenness centrality.

La betweenness centrality di un nodo misura quanto un nodo si trova nel flusso informativo del grafo, supponendo che questo dato flusso informativo segua un percorso minimale. È una misura che fa parte delle misure di centralità di un nodo, e che coglie appunto la centralità di un nodo nello scambio di informazioni all'interno del grafo.

Per calcolarla si considera la frazione di sentieri minimali cui il nodo appartiene, rapportata a tutti gli altri sentieri minimali tra tutte le possibili coppie di nodi del grafo (dove per sentiero minimale si intende il sentiero più breve che collega due nodi).

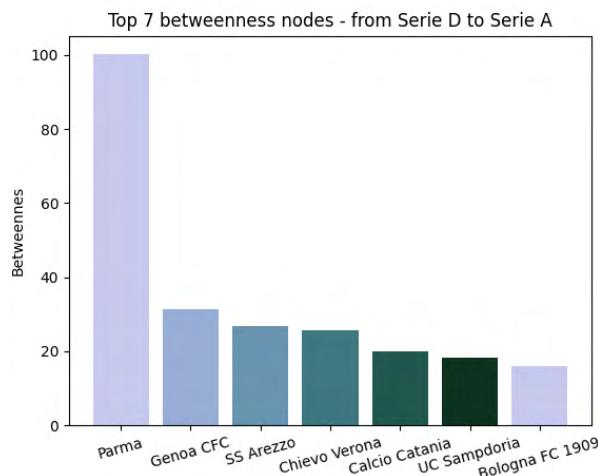
Per lo scopo della nostra analisi non è necessario prendere in considerazione tutti i sentieri minimali di qualsiasi coppia di nodi del grafo, ma i sentieri minimali che collegano i nodi delle squadre di Serie D (e in seguito C e Primavera per il secondo e terzo grafo) con i nodi delle squadre di Serie A.

Si è quindi optato per un calcolo della betweenness che tenesse conto esclusivamente dei sentieri minimali i cui nodi di partenza sono i nodi della Serie D (C, Primavera) e i cui nodi di arrivo sono i nodi della Serie A. Questo è reso possibile dalla libreria networkX di Python che permette di specificare, nel calcolo della betweenness i parametri *sources* e *targets*: nel nostro caso al parametro *sources* verranno passati i nodi della Serie D nel primo grafo, i nodi della Serie C nel secondo grafo, ed i nodi della Primavera nel terzo grafo, mentre al parametro *targets* verranno sempre passati i nodi della Serie A.

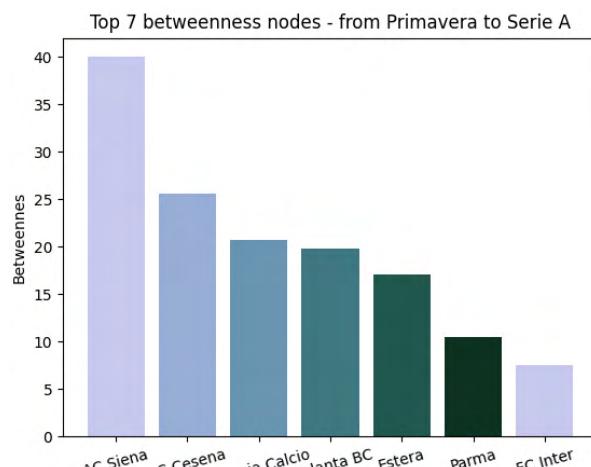
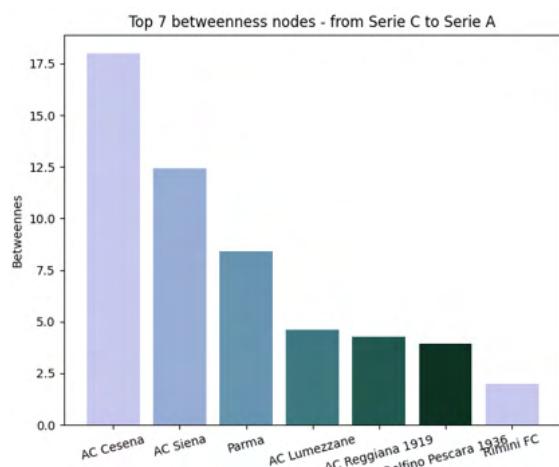
Calcolando in questo modo la betweenness, la supposizione è che:

più la betweenness di un nodo è alta, più quel nodo-squadra sarà presente nei percorsi più brevi tra, le squadre della serie minore presa di volta in volta in considerazione, e le squadre della Serie A, e di conseguenza potrà maggiormente essere considerato come nodo-squadra cruciale per un giocatore per abbreviare il suo percorso verso la serie A.

Come già in parte si poteva intuire dalla rappresentazione dei grafi queste, per ogni grafo, sono le 7 squadre con la più alta betweenness centrality:



Ciò che risulta da questo diagramma a barre è che le top 7 squadre che fanno da ponte verso la serie A per i giocatori che partono dalla serie D appartengono tutte alla serie A ad eccezione dell'Arezzo che è l'unica squadra in serie C.



Da sottolineare la presenza del nodo Estera tra quelli con la betweenness più alta nell'analisi dei percorsi tra squadre Primavera e Serie A.

Significa che per molti giovani calciatori delle squadre Primavera, una delle soluzioni migliori per poi arrivare ad alti livelli nel minor tempo possibile, è quella di migrare verso i campionati esteri.

QUARTO PUNTO

Il quarto ed ultimo punto cerca di rispondere alla domanda se sia possibile stabilire se una squadra è di alto livello, medio livello o basso livello, basandosi soltanto sulle caratteristiche dei calciatori acquistati, provando anche ad individuare quali "skills" portano un calciatore a trasferirsi verso una grande squadra, verso una squadra media, o verso una squadra di basso livello.

Per ogni giocatore abbiamo a disposizione dal dataset diverse informazioni. Per effettuare l'analisi di questo quarto punto abbiamo scelto di prendere in considerazione per i calciatori tre skills ottenibili dal dataset: l'età, il valore di mercato al momento del trasferimento, il campionato di provenienza prima del trasferimento.

Per ognuna di queste 3 skill abbiamo raggruppato in diverse categorie i valori che possono assumere, nello specifico:

- 1) le età sono state raggruppate nelle fasce: 17-22; 23-28; 29-34; 35-40
- 2) i valori di mercato sono stati raggruppati nelle fasce:
 - *low value* (minore di 8 milioni);
 - *medium value* (tra 8 e 16 milioni);
 - *high value* (tra 16 e 24 milioni);
 - *top value* (maggiore di 24 milioni)
- 3) i campionati di appartenenza sono stati raggruppati in:
 - *TOP EU* in cui abbiamo inserito le leghe storicamente più importanti in Europa, ovvero Premier League inglese, Liga spagnola, Bundesliga tedesca, Ligue1 francese;
 - *LOW EU* in cui abbiamo inserito le 20 leghe europee più ricorrenti nel dataset, escludendo quelle presenti in TOP EU;
 - *TOP SA* in cui abbiamo inserito le leghe storicamente più importanti in Sud America, ovvero le massime serie di Brasile, Argentina e Uruguay;
 - *LOW SA* in cui abbiamo inserito le leghe sudamericane non presenti in TOP SA ovvero quelle relative a Cile, Colombia, Paraguay, Bolivia, Perù, Venezuela.

Ci siamo concentrati sul campionato italiano, e come detto appena sopra, abbiamo preso in considerazione le squadre catalogando ciascuna di loro in una delle tre categorie: top club, squadra media, squadra di basso livello.

Per poter catalogare le squadre abbiamo scelto di prendere in considerazione la classifica perpetua della Serie A relativa agli ultimi 15 anni. La classifica perpetua somma i punti ottenuti, anno per anno, delle squadre che hanno partecipato al campionato di Serie A

I dati sulla classifica perpetua sono stati ottenuti anch'essi dal [sito Transfermarkt](#) implementando un codice Python che li estrapolasse in automatico.

La classifica si compone di 40 squadre, che sono le uniche 40 squadre che nella storia hanno partecipato almeno una volta al campionato di Serie A.

Una volta ottenuta questa classifica, abbiamo catalogato le prime 10 di questa classifica come top club, le squadre dalla undicesima alla trentesima posizione come squadre medie, e le squadre dalla trentunesima alla quarantesima posizione come squadre di basso livello, estraendo poi un campione casuale di 20 squadre.

Nel creare il grafo l'idea sottostante è, per ogni trasferimento, prendere in considerazione il nodo della squadra di approdo del calciatore, e creare 3 collegamenti: uno con il nodo dell'età del giocatore trasferito, uno con il nodo del suo market value, e uno con il nodo del campionato di provenienza.

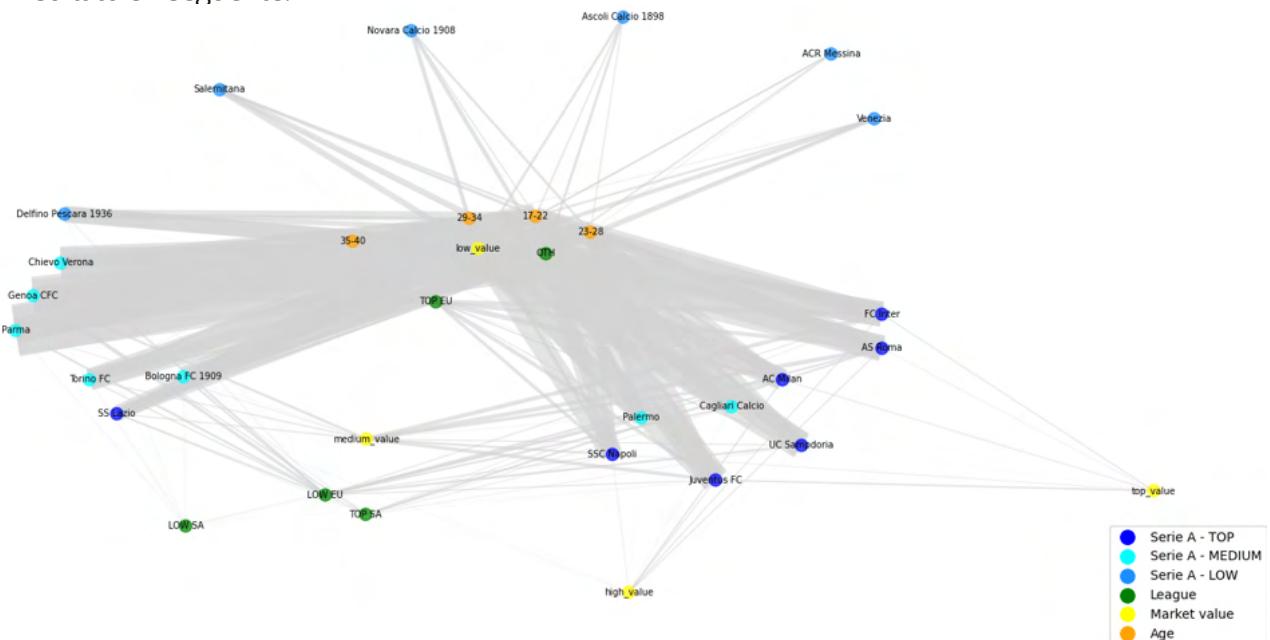
Se per esempio il giocatore Mario Rossi si trasferisce alla Juventus all'età di 24 anni con un valore di mercato di 20 milioni provenendo dalla Premier League inglese, allora il nodo Juventus avrà:

- un arco verso il nodo che rappresenta la fascia di età 20-25
- un arco verso il nodo che rappresenta il valore di mercato 'high value'
- un arco verso il nodo TOP EU che al suo interno ingloba la Premier League inglese come campionato di appartenenza.

Anche in questo caso, come sempre, tutti gli archi che vanno da un nodo ad un altro nella medesima direzione, saranno raggruppati in un unico arco pesato.

La supposizione che ci ha guidato è stata che: se è possibile determinare la fascia di appartenenza di una squadra soltanto dalle caratteristiche relativi ai calciatori acquistati, allora nel grafo i nodi delle skills, 'tirando' i nodi delle squadre, ne avrebbero dovuto influenzare il posizionamento, dando vita alla formazione di cluster di squadre simili per fascia di appartenenza.

Il risultato è il seguente:



È possibile vedere come le squadre (di cui già conosciamo l'etichetta TOP, MEDIUM o LOW grazie alla classifica perpetua) vengano divise quasi alla perfezione dal grafo nei cluster di appartenenza.

Le eccezioni riguardano:

- Palermo e Cagliari, che pur essendo squadre medie nel grafo risultano molto ravvicinate alle squadre top
- La Lazio, per cui vale l'inverso
- Il Pescara, che si avvicina maggiormente alla zona delle squadre medie pur essendo stata pre etichettata come squadra di fascia bassa.

In realtà, nonostante la suddivisione del grafo sia corretta, occorre specificare che probabilmente gran parte del "lavoro" è svolto dalla skill market value, che anche dal grafo sembra essere quella più capace di tirare le squadre da una parte o dall'altra del grafo.

Per quanto riguarda invece le altre due skill, in particolar modo per la skill età, non sembrano essere discriminanti per differenziare le diverse fasce di squadre, e questo probabilmente spiega il perché siano anche le skill più centrali nel grafo, comuni a tutte le squadre indifferentemente.

CONCLUSIONI

Per mezzo della network analysis effettuata abbiamo raggiunto il nostro obiettivo di fornire delle risposte ai quesiti iniziali che ci hanno spinto a scegliere di svolgere questo progetto.

L'analisi svolta inoltre, ci suggerisce come si tratti di una sfera, quella del calciomercato, con un enorme potenziale di dati da sfruttare, e come dunque il nostro lavoro possa essere in futuro anche ulteriormente ampliato.

In primo luogo, il lavoro potrebbe dunque essere esteso non solo ad un'analisi concentrata sull'Italia ma anche sugli altri importanti campionati europei, favorendo anche l'occasione di un migliore confronto tra essi.

In secondo luogo ad essere esteso potrebbe essere il lasso di tempo da prendere in considerazione. Il sito Transfermarkt permette di andare a recuperare i trasferimenti a ritroso senza limite temporale, fino alle date di fondazione delle squadre di calcio.

Un'analisi più estesa dal punto di vista del tempo, potrebbe mostrare come, in un arco temporale superiore ai 15 anni analizzati, si è evoluto il calciomercato e le relative strategie delle squadre sia in Italia che nel resto d'Europa.

BIBLIOGRAFIA

Siti visionati:



Dataset utilizzati:

D2ski, *Football transfers data*, Github, 26 novembre 2021

Emordonez, *Transfermarkt-transfers*, Github, 30 ottobre 2021

Altro materiale:

Greedy modularity communities in Networkx

Girvan-Newman algorithm in Networkx

Schirru F., *I 100 trasferimenti più costosi di sempre*, Goal.Com, 13 agosto 2021

Stephen G. Kobourov, *Spring Embedders and Force Directed Graph Drawing Algorithms*, University of Arizona, 17 gennaio 2012

Chen M., Kuzmin K., *Community Detection via Maximization of Modularity and Its Variants*, IEEE, 2 luglio 2015

Allman A., *13th International Symposium on Process Systems Engineering*, ScienceDirect, 2018

L'Express, *Le PSG officiellement qatari*, Francia, 30 giugno 2011

Brian Skyrms, *Modularity and community structure in networks*, University of California, 19 aprile, 2006