

Analyse des transferts dans les principaux championnats de Football européens

Axel Gardahaut

Résumé

Cet article présente l'étude du marché des transferts européen de football sur la période 2000-2001 à 2018-2019. Dans un premier temps, une étude des caractéristiques du réseau sera présentée, par la suite nous chercherons à étudier les relations entre les différents clubs via les outils d'analyse des réseaux. Différentes approches (pondérations des noeuds et arêtes, graphe dirigé ou non) seront confrontées.

Mots-clés : Social Network Analysis, Community Detection, Football Transfer Market

1 Introduction

Le Football est l'un des sports les plus suivis au monde avec des centaines de millions de fans à travers le monde, générant des milliards de revenus.

L'attrait croissant des communautés américaines et asiatiques jusqu'alors peu engagées boostent ce marché du sport qui contrairement aux sports populaires américains sont en ligues ouvertes sans système de redistribution des talents ainsi un marché de spéculation et inégalitaire s'est progressivement mis en place comme nous ne manquerons pas de l'observer.

Le but final de cette étude est d'observer le marché des transferts via le spectre de l'analyse des réseaux, de comprendre le lien entre la hiérarchie sportive et l'activité économique des clubs, déterminer comment et pourquoi les clubs sont connectés entre eux. Notre étude s'effectue sur l'horizon de 20 ans, ainsi si les dynamiques particulières de certains clubs seront affectées, la structure générale sera relativement stable.

Afin d'éliminer les nombreux très petits clubs apportant de l'information biaisée car leurs transferts n'apparaissent pas dans notre base (trop peu élevés, joueurs issus du centre de formation, prêts), nous nous focaliserons sur les championnats des 5 pays principaux en Europe à savoir la France, l'Angleterre, l'Allemagne, l'Italie et l'Espagne comme une ligue fermée (concentrant 53% de la valeur échangée sur la période).

2 Données

Les données sont obtenues via le site www.transfermarkt.com. On se base sur les transferts les plus importants des années 2000-2001 à la saison 2018-2019. Elles comportent une base de 4200 transferts ainsi que les valeurs marchandes estimées des joueurs à partir de la saison 2006-2007. Des informations complémentaires sur les clubs ont été obtenues par scrapping notamment leurs localisations et leurs participations dans les compétitions européennes qui sont à la fois un proxy de leur compétitivité à l'échelle nationale et européenne.

En effet, pour se qualifier à ces compétitions il faut ou bien finir dans les premières places de son championnat national (le nombre de places octroyées est déterminé par le classement Elo des championnats) ou remporter une coupe nationale (pour les championnats les plus importants). La Ligue des champions étant la compétition des meilleurs clubs européens et la Ligue Europa étant la seconde compétition européenne, la

valeur accordée à la participation n'est pas la même.

En se basant sur les rémunérations des participants à la phase finale des 2 compétitions, la Ligue des champions est 4 fois plus importante (2 MM contre 560 M pour la Ligue Europa en 2020), la mesure de performance issue de ce constat sera notée `elo_eur` dans le code joint.

3 Analyse

Ce réseau peut être considéré comme un graphe orienté (les transferts étant dirigés) ou comme un graphe non orienté (chaque transfert est considéré comme un échange).

On observe une forte disparité tant au niveau économique que sportif pour les différents clubs, sans surprise elles sont fortement positivement corrélées ($\text{cor}=0.87$). Cependant l'efficacité économique (mesurée par les plus values) ne semble pas associée à la performance sportive en soi, les clubs réalisant des plus values ont des scores `elo` non nuls voire élevés cependant il ne s'agit pas de top clubs à fort `elo_score` et performance sportive qui investissent massivement pour acquérir des top joueurs et se rattrapent sur les gains des compétitions et les recettes issues du merchandising mais plutôt de clubs intermédiaires comme Olympique Lyon, Borussia Dortmund, Tottenham, US Palermo.

3.1 Structure du réseau

Le réseau de manière globale semble opérer comme un réseau invariant d'échelle (on trouve $\gamma = 2.5$) cependant au niveau national on retrouve des réseaux plus denses (autour de 10-15% en local comparé à 4% pour le système global) et avec un coefficient de clustering plus élevé.

Une observation des degrés in et out permet d'affiner l'analyse.

Ces réseaux locaux n'en sont pas moins différents, en effet si en Angleterre on semble retrouver un réseau domestique plus proche d'un modèle aléatoire avec une distribution presque normale des degrés, on observe pas ceci dans les autres pays qui ont un marché domestique moins actif que l'Angleterre (ceci s'expliquant en partie par les règles spécifiques du marché anglais (besoin d'un contrat de travail pour jouer et aligner un certain nombre de joueurs anglais par match) et la qualité des joueurs supérieures en moyenne dans les clubs anglais)

s'approchant plus de modèles petit monde avec des $\gamma \leq 2$ tandis qu'on a $\gamma = 3.8$ pour le marché anglais et $\gamma = 3$ pour le marché allemand.

De plus le marché anglais comptabilise plus de clubs susceptibles d'échanger des joueurs comptabilisés dans la mesure où la compétition est plus rude que dans les autres pays ce qui attire plus d'argent via les droits TV et plus de mobilité socialo-sportive des clubs.

Inversement, le championnat français possède moins d'argent via les droits télévisés du fait et en conséquence de la moindre qualité des équipes (somme des scores elo_eur) et reste peu concurrentiel via l'archidomination quasi systématique d'une équipe sur le championnat par conséquent les clubs français exportent beaucoup plus qu'ils n'achètent aux autres championnats considérés.

Les marchés avec $\gamma \leq 2$ possèdent des clubs qui vont capter la grande majorité des transferts. On observe que les clubs étant centraux dans leur marchés nationaux le sont relativement à l'échelle européenne cependant la hiérarchie et les rôles ne sont pas nécessairement les mêmes.

De manière non surprenante, les clubs ayant un fort score performance sont également très important en terme de volume de transferts (centralité de degré).

Il s'agit de clubs ayant une forte valeur de betweenness mais une faible valeur en eigenfactor en conséquence : US Palermo (3ème betweenness, 40ème eigenfactor) ou encore Olympique Lyon qui au niveau national est central mais un intermédiaire de choix pour les tops clubs européens.

Les top clubs échangent beaucoup et également entre eux (45% des transferts sortant d'un top club (classe A) sont vers un autre top club).

Les clubs d'une même catégorie échangent généralement entre eux et avec les catégories directement inférieures et supérieures.

Il n'y a pas de rivalité observable de ce point de vue.

3.2 Détection de communautés

Afin d'étudier les relations, nous avons essayé de détecter des communautés parmi nos clubs. La métrique habituellement utilisée pour évaluer la qualité du partitionnement est la modularité qui permet d'évaluer la différence par rapport à un réseau aléatoire, ie sans structure, ayant le même ordre et la même taille. Il existe plusieurs méthodes de partitionnement, parmi elles, les méthodes divisive et spectrale sont appliquées au graphe. La méthode divisive est basée sur la notion de centralité.

Elle consiste à calculer la centralité d'intermédiation pour tous les liens puis de supprimer le lien ayant la plus forte centralité. Ce processus est répété jusqu'à obtention des communautés. Cette méthode classe les clubs en 46 groupes avec une modularité de 0,22. La méthode spectrale basée sur l'algorithme de Louvain qui classe suivant une approche locale de la modula-

rité : deux sommets sont placés dans une communauté lorsque le gain en modularité est maximal. Ainsi nous obtenons une partition en cinq groupes du graphe avec une modularité de 0,3741.

Ce partitionnement fait ressortir les championnats, et la classification est juste dans la plupart des cas. Par exemple, sur 53 clubs de la Première League, 49 sont regroupés dans une même communauté. Les manquants ne sont pas les moindres puisqu'il s'agit de Chelsea, Man City et Arsenal comptabilisant chacun au moins 1 Milliard d'échanges.

Ils ont été classés avec le championnat espagnol, ce qui peut s'expliquer par leur manque d'implication dans le championnat domestique par rapport aux autres poids lourds du championnat tels que Man United ou Liverpool et leurs échanges avec les poids lourds espagnols : Real Madrid et Fc Barcelone. Cette méthode est celle qui partitionne au mieux le graphe avec la modularité la plus élevée. Nous avons essayé sans succès de prendre en compte l'orientation du graphe pour la classification. La pondération par les effectifs donne des résultats relativement similaires.

4 Conclusion

L'étude aurait été sans doute plus complète en intégrant la totalité des moyens d'échange entre clubs et en prenant en compte le type de joueur échangé.

Un travail sur la hiérarchie sportive et sociale des clubs pourrait être développé pour suivre l'évolution conjointe des clubs et des joueurs.

On pourrait également s'intéresser aux phénomènes de rivalité et domination dans les interactions entre les clubs.

La prise en compte de l'évolution bi-annuelle du réseau pourrait être étudiée, car les performances d'une équipe ne sont pas toujours le reflet de sa place dans la hiérarchie sportive à moyen terme, l'intervention d'investisseurs peut muter cette hiérarchie brutalement, Paris et Man City en sont les parfaites illustrations.

Références

- [1] FORTUNATO.S : *Community detection in graphs*
- [2] PASCAL EUSEBIO, DAVID LEVY, JEAN MICHEL FLOCH : *Partitionnement et analyse de graphes*
- [3] BIERMAN.C : *Big Data Foot*
- [4] LAURENT BEAUGUITTE, CÉSAR DUCRUET : *Scale-free, small-world networks et géographie*
- [5] HAMED SEYED-ALLAEI AL : *Scale-free networks with an exponent less than two* The Abdus Salam International Center for Theoretical Physics

A Graphe du marché européen

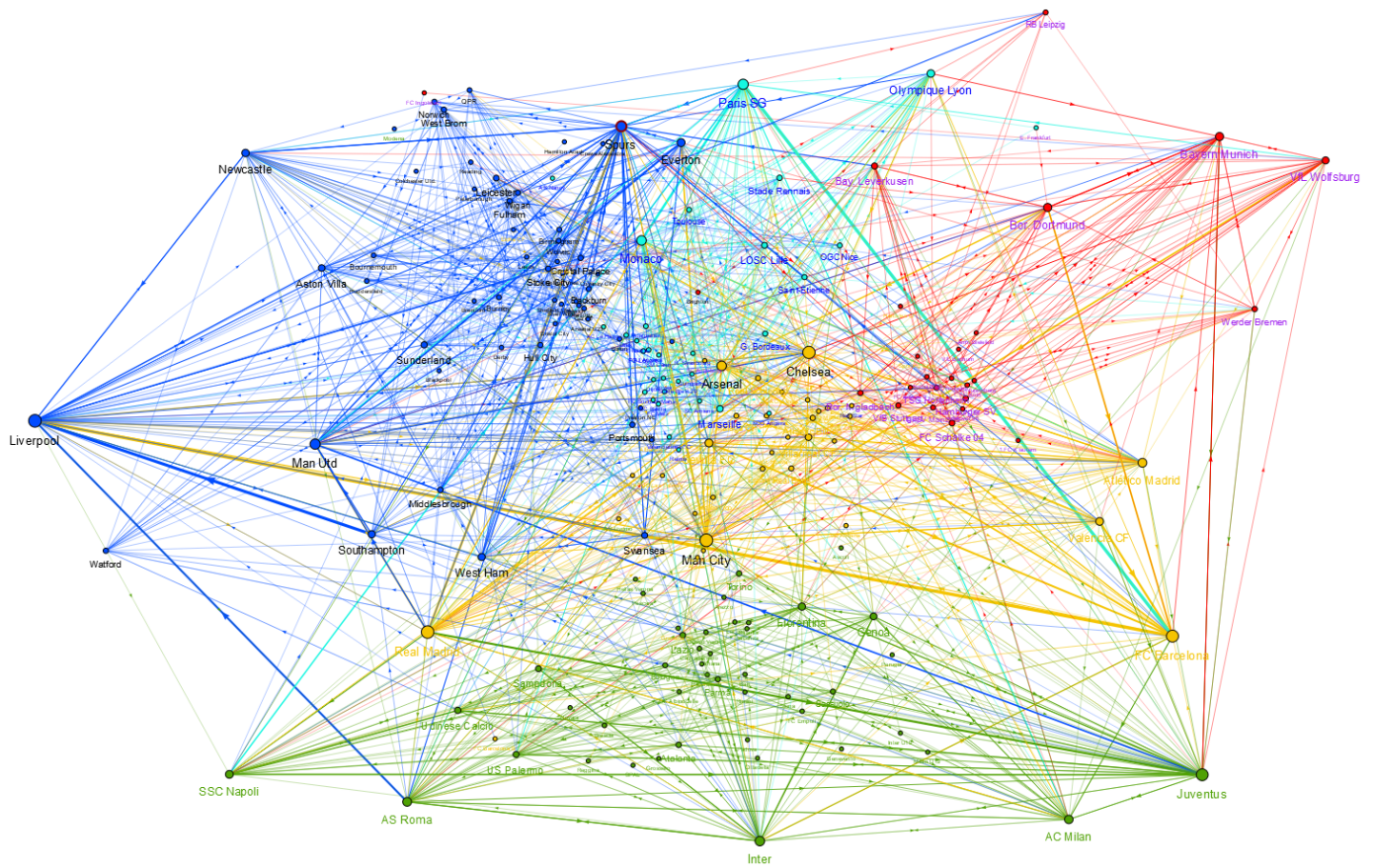


FIGURE 1 – Clusters obtenus via Louvain : Cyan : France, Rouge : Allemagne, Jaune : Espagne, Bleu : Angleterre, Vert : Italie