

A Social Network Analysis of Football with Complex Networks

Lucas G. S. Félix

Federal University of São João del Rei
São João del Rei, Minas Gerais
lucasgsfelix@gmail.com

Vinícius da F. Vieira

Federal University of São João del Rei
São João del Rei, Minas Gerais
vinicius@ufsj.edu.br

Carlos M. Barbosa

Federal University of São João del Rei
São João del Rei, Minas Gerais
cmagnobarbosa@gmail.com

Carolina Ribeiro Xavier

Federal University of São João del Rei
São João del Rei, Minas Gerais
carolinaxavier@ufsj.edu.br

ABSTRACT

Soccer is the most popular sport in the world and due its popularity, soccer moves billions of euros over the years, in most diverse forms, such as marketing, merchandising, TV quotas and players transfers. As example, in the 2016/2017 season, only England has moved about 1.3 billion of euros only in players transfers. In this work, it is performed a study of the transfer market of player. To do so, players transfer data were gathered from the website Transfermarkt and were modeled as a graph. In order to perform this study, different Complex Networks techniques were applied, such as Overlap Community Detection and Property Analysis. Through our results we could evaluate the soccer players market, and see a pattern that every market has at least one farm country, which has a main function of selling athletes, or a buyer country, which most of its transactions is buying players.

KEYWORDS

Complex Networks, Data Mining, Social Network, Sport Analytics

1 INTRODUÇÃO

Atualmente, com o crescente volume e variabilidade de dados, a tarefa de interpretar e transformar estes dados em informações tornou-se difícil e trabalhosa. Contudo, a mesma torna-se essencial para que as informações geradas por grandes dados nos faça ter uma percepção melhor, mais coesa e exata sobre sistemas e usuários, fazendo com esta seja uma área de estudo de interesse de empresas e pesquisadores [10].

Uma das formas que nos permite ter um melhor entendimento dos dados são as análises sociais. Por meio deste tipo de análise é possível ter uma melhor compreensão do comportamento de ligações complexas, tendo seus princípios na sociologia e tornando-se um problema computacional por meio de áreas de estudo como redes complexas [3]. As redes complexas por sua vez, modelam problemas por meio de um grafo, permitindo uma representação mais fidedigna de interações sociais, tornando este tipo de abordagem ideal para modelar ligações sociais como redes de coautoria e amizades [2].

A análise de dados massivos para avaliação de esportes tem sido cada vez mais comum. Empresas auxiliam times e esportistas por

meio de cruzamento de dados, previsão de padrões para montagem de times competitivos, avaliação de adversários e análise do time em tempo real [1], mostrando que a informação rápida e atual é um diferencial para o time [4]. Na literatura atual diversos estudos propõe análises do esporte, seja do ponto de vista computacional [7, 8, 18, 20, 21], utilizando dados para avaliação, da educação física [12, 14, 15] ou do sociológico [9, 13, 16, 19].

Levando em consideração estes aspectos, este trabalho visa a utilização de técnicas de redes complexas e análises sociais para avaliação de uma rede de transferências de jogadores de futebol entre países. A escolha deste tema se dá pelo fato do futebol ser na atualidade o esporte mais popular do mundo [13, 16]. Assim, o futebol gera um enorme fluxo financeiro vindo de *marketing*, *merchandising*, venda de ingressos e transações de jogadores, possuindo uma movimentação de 1,3 bilhões de euros na temporada 2016/2017 somente na Inglaterra. A avaliação das transferências de jogadores de futebol se dá pois pela importância das transações de atletas para clubes de futebol. Além de ser uma das maneiras mais efetivas de compensação monetária que um clube pode possuir [13], as vendas e compras de jogadores mostram o fluxo de trabalho entre países e como estes se comportam em um mercado de transferências [8].

Para realização deste trabalho foi desenvolvido um algoritmo de coleta e parser de dados, o qual coletou dados de transferências de jogadores de futebol do site Transfermarkt.¹ Após a coleta dos dados, foi realizada a modelagem da rede, sendo feito um estudo e caracterização do grafo gerado. Para uma avaliação mais detalhada da rede, foi utilizado o algoritmo de detecção de comunidades com sobreposição proposto em [11].

2 TRABALHOS RELACIONADOS

Na literatura, existem atualmente diversos estudos que abordam o futebol como tema principal, com pesquisas que se estendem desde a educação física a sociologia. Baseado em pesquisas realizadas até o presente momento, são avaliados aqui trabalhos que se relacionam com este em aspectos metodológicos, com a utilização de algoritmos e técnicas de redes complexas, e em relação ao tema futebol.

A análise de redes consiste na tarefa de avaliação de grafos utilizando algoritmos de redes complexas, sendo esta amplamente aplicada em tempos atuais para o estudo de sistemas e ligações entre objetos complexos. Trabalhos como [5, 6], utilizam de técnicas de redes complexas para avaliação de times e jogadas no futebol. Em [6] é avaliada a seleção da Espanha na copa do mundo de 2010. Neste

In: XVI Workshop de Trabalhos de Iniciação Científica (WTIC 2019), Rio de Janeiro, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (Web-Media). Porto Alegre: Sociedade Brasileira de Computação, 2019.
ISSN 2596-1683

¹transfermarkt.com

trabalho, os jogadores são nós do grafo, e os passes que os jogadores fazem são arestas direcionadas, sendo feitas análises temporais no grafo. Por meio desta metodologia, os autores conseguiram avaliar que jogadas coletivas eram o ponto forte do time, possuindo como nós mais centrais, jogadores de maior habilidade.

No trabalho [5] os autores avaliam posições que mais auxiliam em jogadas de ataque utilizando dados da copa do mundo de 2014. No trabalho foram avaliados todos os times e partidas da competição. Por meio da metodologia proposta, utilizando diferentes métricas de centralidade de rede, os autores identificaram a posição de meio de campo como principal posição para construir jogadas.

Dentre os diversos trabalhos que estudam o futebol na literatura, poucos avaliam diretamente as transferências de atletas do esporte estudado. Em [17], os autores dão uma visão de otimização para ao problema, considerando a melhor maneira de se construir um time a partir de transferências. Assim, um modelo de programação linear que visa maximizar as habilidades de atletas e diminuir o custo de jogadores é proposto. Por meio da metodologia, é possível definir melhores jogadores a serem contratados de acordos com posições de maior carência em um time de futebol.

Os artigos [8, 21], bem como este, utilizam-se de métricas de rede para avaliação do mercado de transações de atletas. O trabalho [21], tem como principal objetivo analisar o sucesso de um time de acordo com suas transações. Para isso os autores utilizam-se de dados de transferências de 2011 – 2015, onde os nós são times e as ligações são direcionadas do time vendedor para o time comprador. Por meio das avaliações propostas, os autores concluíram que apesar do dinheiro auxiliar na montagem de um time consistente, é possível vencer campeonatos evitando gastos exorbitantes. Em [8], os autores avaliam as transferências feitas entre os países presentes na copa do mundo de 2018. Considerando uma modelagem onde os países são os nós e as arestas são transações entre países, os autores concluem que há uma distribuição de calda longa na rede, mostrando que poucos países concentram grande maioria das transferências realizadas. Além disso, o trabalho mostra que grande parte dos países considerados vendedores, são países que possuem uma economia mais fraca, sendo o futebol um dos escapes para a população.

Este trabalho possui como diferencial o fato de sua metodologia utilizar uma abordagem de comunidades com sobreposição. Utilizando este método é possível avaliar o mercado de transações de maneira mais dinâmica, não limitando cada país a apenas um grupo, como ocorre em [8], onde os autores utilizando uma abordagem sem sobreposição para detecção de comunidades.

3 REDES COMPLEXAS

Diversos sistemas na natureza e no dia a dia são considerados sistemas complexos. A interação entre satélites que permite a comunicação global, a integração entre neurônios, a internet e a própria economia global são exemplos de grandes sistemas complexos. A interação entre milhares de componentes recebe a nomenclatura de sistemas complexos. A própria organização dos estados modernos é outro grande exemplo de sistema complexo, com bilhões de indivíduos cooperando para o desenvolvimento de países. Por trás de cada um destes sistemas complexos existe um grafo, que

permite a interação entre milhares de nós formando assim uma rede complexa.

Dentro do tópico de redes complexas, um dos conceitos com grande importância são as comunidades ou grupos. A detecção de comunidades é dos principais tópicos de estudo dentro de redes complexas e a detecção de tais comunidades se apresenta como um problema desafiador que é atacado de diversas maneiras. Este conceito se refere ao objetivo de encontrar agrupamentos de nós que possuem a maioria das ligações com outros nós que estão neste mesmo grupo do que com elementos que estão em outros grupos.

Os métodos de detecção de comunidades na sua grande maioria realizam a identificação de agrupamento através de um processo de particionamento dos vértices. No qual cada vértice pertence a uma única comunidade. Porém em grande parte dos cenários reais, como nas redes sociais, os usuários possuem múltiplos interesses, podendo integrar mais de uma comunidade simultaneamente. Com este intuito a identificação de comunidades com sobreposição são de grande interesse de estudo da comunidade científica.

Uma abordagem para encontrar estas comunidades sobrepostas é uma técnica apresentada por Lancichinetti, na qual é utilizada uma estratégia de otimização local no qual uma função de *fitness* é maximizada. Este método considera que as estruturas de comunidades são essencialmente locais e se busca encontrar as comunidades formadas em torno de determinados nós. A escolha deste nós são definidas pelo algoritmo de maneira aleatória. O processo de detecção das comunidades pode ser resumido nas etapas abaixo:

- (1) Escolha um nó A;
- (2) Detecte a comunidade natural em torno deste nó;
- (3) Escolha de maneira aleatória um nó B que não pertence a nenhum grupo;
- (4) Detecte a comunidade natural do vértice B, explorando todos os possíveis membros inclusive os que já pertencem a uma comunidade;
- (5) Repita do passo 3.

O processo de detecção da comunidade natural ocorre através da maximização da função de *fitness*, na qual os vizinhos da comunidade são avaliados e escolhido os vizinhos que propiciam uma melhoria no índice da comunidade. A Equação 2.2 representa a função de *fitness* que será maximizada em uma comunidade.

$$fG = \frac{k_{in}^G}{(k_{in}^G + k_{out}^G)^\alpha} \quad (1)$$

Onde o k_{in}^G representa o total das ligações internas e k_{out}^G o total das ligações externas dos nós na comunidade, e α é um número real positivo que controla o tamanho do agrupamento.

4 METODOLOGIA

Este trabalho propõe uma análise das transferências de jogadores de futebol por meio de técnicas de redes complexas. Para realização de tal, é proposta uma abordagem em três vias para realização das análises, sendo estas (i) Coleta de dados, (ii) Modelagem dos dados, e (iii) Análise das redes.

4.1 Coleta de dados

Para realização de nosso estudo, fora realizada a coleta de dados sobre o site Transfermarkt². O site consiste em um grande banco de dados com diversas informações relacionadas ao futebol como estatísticas, tabelas de campeonatos e, o principal objeto de estudo deste trabalho, os dados relacionados a transferências de jogadores. Para a realização de tal tarefa foi desenvolvido por um *crawler* e junto a este um *parser*, com objetivo de realizar uma limpeza e um pré processamento dos dados. Ao final os dados foram armazenados em um banco de dados, ficando em formato estruturado.

Para a coleta foi considerado o período de 1962 até 2017, entretanto, foi gerada uma única rede que compreende as transferências realizadas entre os anos de 1990 à 2018. A escolha do período deu-se pela maior quantidade de transferências realizadas durante este período, sendo os valores anteriores ao ano de 1990 muito pequenos para análises aprofundadas.

4.2 Modelagem da rede

Seja $S = \{s_1, \dots, s_n\}$ a lista de seleções consideradas. A rede de transferências de jogadores foi modelada como um grafo $G = (V, A)$, onde cada vértice $v_i \in V$ representa um país $s_i \in S$ e cada arco $a_{ij} \in A$ representa a relação de venda de um jogador do país s_i para o país s_j . Além disto, uma função $f : a_{ij} \mapsto p_{ij}$ associa cada arco $a_{ij} \in A$ com um peso $p_{ij} \geq 0$ que representa o número de vendas de jogadores realizadas do país s_i para o país s_j , caso exista transações naquela direção. Ao final a rede gerada possui 94 vértices e 1898 arestas.

4.3 Análise da rede

Para realização da análise da rede, foi feita uma caracterização da mesma através de suas propriedades. Por meio da caracterização é possível descobrir atributos importantes da rede que nos auxiliam em um melhor entendimento do grafo. Também foi realizada a análise de comunidades da rede gerada. Para isso, foi utilizado o algoritmo proposto por Fortunato [11].

Em [11], os autores propõem um método para detecção de comunidades com sobreposição. Diferente, da detecção de comunidades tradicional, em que um elemento obrigatoriamente pertence a apenas um grupo, em algoritmos que consideram sobreposição, elementos podem fazer parte de um ou mais grupos, permitindo assim, uma melhor caracterização da rede [22].

5 RESULTADOS

Nesta seção são apresentados os resultados obtidos através de nossa metodologia. Inicialmente foi feita uma caracterização da rede através de suas propriedades. Em uma segunda etapa, foi realizada a detecção de comunidades com sobreposição sobre a rede, revelando grupos comerciais coesos entre os países avaliados.

5.1 Caracterização da rede

Para realização da caracterização da rede foram avaliadas as propriedades da rede. As propriedades de uma rede nos dão informações importantes, que corroboram com a realidade, e revelam

novas informações para análises mais aprofundadas. Na Tabela 1, é possível ver as principais propriedades da rede.

Propriedade	Valor
Densidade	0.21
Reciprocidade	0.56
Assortatividade grau	-0.09
Grau máximo entrada	66
Grau máximo saída	50
Diâmetro	7.0

Table 1: Propriedades da rede de transferências de jogadores

A densidade da rede é a razão entre a quantidade de arestas do grafo e a quantidade de arestas do grafo completo com o mesmo número de vértices, desta forma, quanto mais perto de um, mais ligados são os vértices. Em nossa rede, essa medida ficou próxima a 0.21, mostrando que a maioria dos países não fazem transações entre si. Isso é corroborado pela medida de reciprocidade, que é a medida da possibilidade de um par de vértices de um grafo direcionado serem mutuamente ligados, o valor apresentado na Tabela 1 informa que a rede é pouco mútua. Isso nos mostra que se $X \rightarrow Y$, ou seja, X vende um jogador para Y , a probabilidade de $X \leftarrow Y$, ou seja, X comprar um atleta de Y é baixa.

A assortatividade de grau mostra a tendência de um vértice ter uma ligação com outros vértices de grau similar ao dele. Na rede estudada, este valor está próximo de zero, mostrando que a rede não é assortativa e que países não tem uma preferência de ligações com países de graus altos ou baixos.

As métricas de grau máximo de entrada e saída, mostram os vértices que fazem transações de saída (venda de jogadores) com o maior número de países e o país que faz transações de entrada (compra de jogadores) com o maior número de países, sendo estes Brasil e Rússia respectivamente.

Vale ressaltar que estes fatos corroboram com o trabalho de [8], já que o Brasil possui uma balança comercial extremamente positiva, com o maior número de vendas. Já a Rússia, possui uma balança extremamente negativa, possuindo várias transferências de compra, evidenciado assim a existência de mercados compradores e mercados fazenda.

O diâmetro é a maior entre os menores caminhos entre todos os pares de vértices. E pode ser interpretado como o número máximo de transferências que um jogador teria que fazer para sair de um país A e chegar em um país B . Isso mostra que, apesar de muitos países estarem a uma distância curta de outros, existem mercados pouco explorados. Dada a dinâmica atual do mercado de futebol, os jogadores precisariam de vários intermediários para sair do país A e passar a atuar no país B ou vice-versa.

5.2 Análise de Comunidades

A detecção de comunidades auxilia no encontro de grupos coesos, nos quais os elementos pertencentes a um grupo são mais semelhantes quando comparado a elementos fora desse grupo.

Por meio da utilização da técnica proposta em [11], foi possível identificar mercados onde países costumam realizar a grande maioria de suas transações, sendo identificadas 11 comunidades. Muitas dessas comunidades são formadas por países que se situam no mesmo continente ou que estão próximos geograficamente, sendo

²transfermarkt.com

estes resultados mais evidentes. Além disso, uma comunidade gigante foi encontrada, a qual possui todos os vértices do grafo. Devido ao escopo reduzido deste trabalho, iremos de apresentar de maneira mais detalhada apenas algumas comunidades.

Assim, em nossas avaliações destacamos duas comunidades. O primeiro grupo é formado por Colômbia, Peru, Argentina, Paraguai, Chile, Bolívia, Equador, México e Malásia. Essa comunidade se caracteriza por mercados fazenda [21], com exceção do México. Países fazenda, são caracterizados pela sua grande quantidade de vendas, realizando poucas transações de compras. Assim, percebe-se que essa comunidade é formada por um grande país comprador, México, e países fazenda que vendem para tal. Vale, ressaltar que alguns destes países, como México, Argentina e Peru, estão presentes em mais de uma comunidade, mostrando assim que há uma variação em seus mercados de compra e venda. Nesta comunidade, o grande país vendedor é a Argentina, a qual é o segundo país que mais possui transações de venda, ficando atrás apenas do Brasil [8].

A segunda comunidade avaliada é formada por países como Brasil, Egito, Líbano, Vietnã, Tailândia, Emirados Árabes e Arábia Saudita. Essa comunidade, se diferencia da primeira por possuir um grande país vendedor ligado a vários outros compradores, que buscam fortalecer seu mercado nacional. O Brasil, como grande vendedor, se mostra o principal país considerado fazenda [8], fornecendo jogadores a países que possuem campeonatos internos mais fracos, contudo, estão dispostos a investir em jogadores estrangeiros para que haja uma consolidação de suas ligas.

O restante das comunidades estão presentes na tabela disponibilizada no link ³. Nesta é possível ver todos os países em cada comunidade.

6 CONCLUSÃO

Neste trabalho, é proposto um estudo de metodologias de redes complexas com estudo de caso sobre uma rede de transferências de jogadores de futebol. Neste é dado um foco na importância das propriedades do grafo e como estes podem nos auxiliar em nossas avaliações. Além disso, é feita uma avaliação utilizando um algoritmo de detecção de comunidades com sobreposição, para análise de possíveis mercados existentes.

Por meio de meio de nossas análises podemos perceber que países de maior poder econômico e países fazenda são países que possuem diferentes papéis na rede, mas que realizam a maior quantidade de transferências. Isso influencia diretamente nas comunidades encontradas em nossa avaliação, dado que as mesmas podem possuir duas configurações: (i) Conexão entre um país vendedor com diversos países compradores (ii) Conexão de diversos países vendedores um país comprador.

Em trabalhos futuros, pretende-se avaliar de maneira mais aprofundada a configuração das comunidades, comparando diferentes tipos de algoritmos de detecção de comunidades com sobreposição. Isso é guiado pela grande quantidade de países presentes em apenas uma comunidade, sendo boa parte destes países importantes para balança comercial mundial fazendo o papel de grandes consumidores como França, Inglaterra, Espanha, Alemanha e Itália. Além

disso, pretende-se realizar análises de centralidade, para a identificação dos principais países em nossa rede e análises de predição de links, avaliando possíveis mercados que podem se ligar.

REFERENCES

- [1] [n.d.]. FORBES, The Australian Tech That's Improving The World's Best Athletes. <https://www.forbes.com/sites/alexkonrad/2013/05/08/austrie-tech-catapult-gps/#3097b434c02a>. Accessed: 2019-06-06.
- [2] Victória Patrícia Aires, Fabiola G Nakamura, and Eduardo F Nakamura. 2017. A Social Network Analysis of Successful Movies with Complex Networks. In *Anais Estendidos do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web*. SBC, 99–102.
- [3] Albert-László Barabási et al. 2016. *Network science*. Cambridge university press.
- [4] Christopher Carling, Jonathan Bloomfield, Lee Nelsen, and Thomas Reilly. 2008. The role of motion analysis in elite soccer. *Sports medicine* 38, 10 (2008), 839–862.
- [5] Filipe Manuel Clemente, Fernando Manuel Lourenço Martins, P Del Wong, Dimitris Kalamaras, and Rui Sousa Mendes. 2015. Midfielder as the prominent participant in the building attack: A network analysis of national teams in FIFA World Cup 2014. *International Journal of Performance Analysis in Sport* 15, 2 (2015), 704–722.
- [6] Carlos Cotta, Antonio M Mora, Juan Julián Merelo, and Cecilia Merelo-Molina. 2013. A network analysis of the 2010 FIFA world cup champion team play. *Journal of Systems Science and Complexity* 26, 1 (2013), 21–42.
- [7] Leonardo Cotta, POV de Melo, Fabricio Benevenuto, and Antonio AF Loureiro. 2016. Using fifa soccer video game data for soccer analytics. In *Workshop on Large Scale Sports Analytics*.
- [8] Lucas GS Félix, Carlos M Barbosa, Iago A Carvalho, Vinícius da F Vieira, and Carolina Ribeiro Xavier. 2018. Uma análise das seleções da copa utilizando uma rede de transferências de jogadores entre países. In *7º Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2018)*. SBC.
- [9] Grant Jarvie. 2003. Sport, racism and British society: A sociological study of England's elite male Afro/Caribbean soccer and rugby union players. In *Sport, racism and ethnicity*. Routledge, 79–102.
- [10] Felix L. G. S., Caetano J. V., C. R., W. Luiz, Dias D., and Rocha L. D. 2018. Avaliação Automática de Conteúdo de Aplicações de Reclamação Online. *The Symposium on Knowledge Discovery, Mining and Learning (KDMiLe)* 6 (2018).
- [11] Andrea Lancichinetti, Santo Fortunato, and Janos Kertesz. 2009. Detecting the overlapping and hierarchical community structure in complex networks. *New journal of physics* 11, 3 (2009), 033015.
- [12] Adrian Lees, Takeshi Asai, T Bull Andersen, Hiroyuki Nunome, and Thorsten Sterzing. 2010. The biomechanics of kicking in soccer: A review. *Journal of sports sciences* 28, 8 (2010), 805–817.
- [13] Jorg Liebig, Alexander Von Rhein, Christian Kastner, Sven Apel, Jens Dorre, and Christian Lengauer. 2012. Large-Scale Variability-Aware Type Checking and Dataflow Analysis. (2012).
- [14] Jonathan Magee and John Sugden. 2002. "The World at their Feet" Professional Football and International Labor Migration. *Journal of sport and social issues* 26, 4 (2002), 421–437.
- [15] Cristian Osgnach, Stefano Poser, Riccardo Bernardini, Roberto Rinaldo, and Pietro Enrico Di Prampero. 2010. Energy cost and metabolic power in elite soccer: a new match analysis approach. *Med Sci Sports Exerc* 42, 1 (2010), 170–178.
- [16] Ignacio Palacios-Huerta. 2004. Structural changes during a century of the world's most popular sport. *Statistical Methods and Applications* 13, 2 (2004), 241–258.
- [17] Vineet M Payyappalli and Jun Zhuang. 2019. A data-driven integer programming model for soccer clubs' decision making on player transfers. *Environment Systems and Decisions* (2019), 1–16.
- [18] Leandro AA Silva, Johnnatan Messias, Mirella M Moro, Pedro Olmo Vaz de Melo, and Fabricio Benevenuto. 2015. Algoritmos de Aprendizado de Máquina para Predição de Resultados das Lutas de MMA. (2015).
- [19] Ian Taylor. 2014. On the sports violence question: soccer hooliganism revisited. *Sport, Culture and Ideology (RLE Sports Studies)* (2014), 152.
- [20] Pedro OS Vaz de Melo, Virgílio AF Almeida, Antonio AF Loureiro, and Christos Faloutsos. 2012. Forecasting in the NBA and other team sports: Network effects in action. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6, 3 (2012), 13.
- [21] Liu XF, Liu Y-L, Lu X-H, Wang Q-X, and Wang T-X. 2016. The Anatomy of the Global Football Player Transfer Network: Club Functionalities versus Network Properties. *PLoS ONE* 11, 6 (05 Feb 2016). <https://doi.org/10.1371/journal.pone.0156504>
- [22] Jierui Xie, Stephen Kelley, and Boleslaw K Szymanski. 2013. Overlapping community detection in networks: The state-of-the-art and comparative study. *Acm computing surveys (csur)* 45, 4 (2013), 43.

³https://drive.google.com/file/d/12WQSPUUVc50thJRztupz20QGH_mqVyQU/view?usp=sharing