ISCTE IUL
Instituto Universitário de Lisboa
Departamento de Ciências e Tecnologias da Informação

FACULDADE · DE · CIÊNCIAS UNIVERSIDADE · DE · LISBOA
Departamento de Informática

Complex Networks Analysis in Team Sports Performance: Multilevel Hypernetworks Approach to Soccer Matches

A Thesis presented in partial fulfillment of the Requirements
for the Degree of *Doctor of Philosophy* in
Complexity Sciences

By
João Paulo Duarte Ramos

Thesis advisor: Rui Jorge Henriques Calado Lopes
Thesis advisor: Duarte Fernando da Rosa Belo Patronilho de Araújo

July 2019

ISCTE IUL
Instituto Universitário de Lisboa
Departamento de Ciências e Tecnologias da Informação

FACULDADE · DE · CIÊNCIAS UNIVERSIDADE · DE · LISBOA
Departamento de Informática

Complex Networks Analysis in Team Sports Performance: Multilevel Hypernetworks Approach to Soccer Matches

A Thesis presented in partial fulfillment of the Requirements
for the Degree of *Doctor of Philosophy* in
Complexity Sciences

By
João Paulo Duarte Ramos

Jury Members:
Doctor Rui J. Lopes, Assistant Professor, Instituto Universitário de Lisboa
Doctor Duarte Araújo, Professor Associado com Agregação, Universidade de Lisboa
Doctor Keith Davids, Full Professor, Sheffield Hallam University
Doctor Anna Volossovitch, Professora Auxiliar, Universidade de Lisboa
Doctor Jorge Louçã, Professor Auxiliar com Agregação, Instituto Universitário de Lisboa
Doctor Jeffrey Johnson, Full Professor, Open University (UK)

President of the Jury:
Doctor Carlos Sá da Costa, Professor Associado, Instituto Universitário de Lisboa

July 2019

To my wife Ana and my daughters Leonor and Carolina

Abstract

Humans need to interact socially with others and the environment. These interactions lead to complex systems that elude naïve and casuistic tools for understand these explanations. One way is to search for mechanisms and patterns of behavior in our activities. In this thesis, we focused on players' interactions in team sports performance and how using complex systems tools, notably complex networks theory and tools, can contribute to Performance Analysis. We began by exploring Network Theory, specifically Social Network Analysis (SNA), first applied to Volleyball (experimental study) and then on soccer (2014 World Cup). The achievements with SNA proved limited in relevant scenarios (e.g., dynamics of networks on *n-ary* interactions) and we moved to other theories and tools from complex networks in order to tap into the dynamics on/off networks. In our state-of-the-art and review paper we took an important step to move from SNA to Complex Networks Analysis theories and tools, such as Hypernetworks Theory and their structural Multilevel analysis. The method paper explored the Multilevel Hypernetworks Approach to Performance Analysis in soccer matches (English Premier League 2010-11) considering *n-ary* cooperation and competition interactions between sets of players in different levels of analysis. We presented at an international conference the mathematical formalisms that can express the players' relationships and the statistical distributions of the occurrence of the sets and their ranks, identifying power law statistical distributions regularities and design (found in some particular exceptions), influenced by coaches' pre-match arrangement and soccer rules.

Keywords: Complex Systems; Social Networks Analysis; Multilevel Hypernetworks Approach; Performance Analysis; Team Sports; Soccer.

Resumo

Os humanos necessitam interagir socialmente com os outros e com o envolvimento. Essas interações estão na origem de sistemas complexos cujo entendimento não é captado através de ferramentas ingénuas e casuísticas. Uma forma será procurar mecanismos e padrões de comportamento nas atividades. Nesta tese, o foco centra-se na utilização de ferramentas dos sistemas complexos, particularmente no contributo da teoria e ferramentas de redes complexas, na Análise do Desempenho Desportivo baseado nas interações dos jogadores de equipas desportivas. Começámos por explorar a Teoria das Redes, especificamente a Análise de Redes Sociais (ARS) no Voleibol (estudo experimental) e depois no futebol (Campeonato do Mundo de 2014). As aplicações da ARS mostraram-se limitadas (por exemplo, na dinâmica das redes em interações *n-árias*) o que nos trouxe a outras teorias e ferramentas das redes complexas. No capítulo do estado-da-arte e artigo de revisão publicado, abordámos as vantagens de utilização de outras teorias e ferramentas, como a análise Multinível e Teoria das Híperredes. No artigo de métodos, apresentámos a Abordagem de Híperredes Multinível na Análise do Desempenho em jogos de futebol (*Premier League* Inglesa 2010-11) considerando as interações de cooperação e competição nos conjuntos de jogadores, em diferentes níveis de análise. Numa conferência internacional, apresentámos os formalismos matemáticos que podem expressar as relações dos jogadores e as distribuições estatísticas da ocorrência dos conjuntos e a sua ordem, identificando regularidades de distribuições estatísticas de *power law* e *design* (encontrado nalgumas exceções estatísticas específicas), promovidas pelos treinadores na preparação dos jogos e constrangidas pelas regras do futebol.

Palavras chave: Sistemas Complexos; Análise Social de Redes; Abordagem de Híperredes Multinível; Análise do Desempenho Desportivo; Equipas Desportivas; Futebol.

Acknowledgements

This long doctoral process goes far beyond the list of publications, surpassing the accumulated knowledge and scientific experience. The collaboration and support from many people that surrounded me for more than 7 years were notably encouraging and rich. The contributions are not always obvious, and sometimes, little comments from dearest friends that ask, what and how was I doing, had the power to clarify or make me change some important details or directions; some other times, when patient people continuously supported my baby steps improvement. I am grateful for all those who helped me during these years, even though; my exhausted memory will possibly forget some names over here.

My wife, Ana, deserves a special spot. More than love and moral support, she was always the compass regarding my time management and priorities, sacrificing herself with daughters' education, domestic tasks, and especially with my mental and physical "absence". The patience and love of my two daughters, Leonor and Carolina that also suffered with a "workaholic" dad and remained resilient in their own progress and life. My mother, Georgina, with love and calming attitude had also supported me many times overcoming the long distance that separate us, along with my brother, Francisco and sister Ana.

Certainly Rui Jorge Lopes was a key player in this story, his encouragement, prioritizing and inspiring scientific attitude guided and provided me priceless experiences during long hours in these years. On the other hand, not always much time means quality and possessing a visionary experience in publishing and advising, Duarte Araújo, has the ability to transform fatigued mind and ideas in genius ones. They have provided me invaluable lessons about living the science and the research practices.

My friend Pedro Passos in our long bicycle cross country rides and in the International Scientific Congress have pointed me many significant advices and directions on the work to be done. Similarly, Luis Vilar, had contributed via Professional and Academic job opportunity. The feedbacks from Jorge Louçã, Helder Coelho and Jeff Johnson contributed to improve the quality of my work and should be remembered.

Good lunch times and coffee-breaks were enjoyable with many other people in funny and productive moments that are unforgettable with all those people I met during the scientific conferences, workshops and classes; I simply have to skip all those names here.

I cannot forget my friends Bastos' in several lunches and dinners, where scientific and non-scientific conversations toke place.
After all, my daily life in my main work place in the two public schools that I have work in, gave me some closer friends that participated in my life in different ways. Apart from some names already cited, thanks to, Paulo Caldeira, Orlando Fernandes and Sofia Fonseca, which at different time helped me in valuable advice.
I should also thank ISCTE-IUL and CIPER/FMH-UL for supporting my participation, in Complex Systems in Sports - International Congress, in Barcelona (2017) and on the papers published.

Publications

The thesis is based on the following publications (reprinted with the permission of the publishers):

Core Publications (ISI indexed journals):

I. Ramos, J., Lopes, R. J., & Araújo, D. (2017). What's Next in Complex Networks? Capturing the Concept of Attacking Play in Invasive Team Sports. *Sports Medicine*. doi:10.1007/s40279-017-0786-z

II. Ramos, J., Lopes, R. J., Marques, P., & Araújo, D. (2017). Hypernetworks Reveal Compound Variables That Capture Cooperative and Competitive Interactions in a Soccer Match. *Frontiers in Psychology, 8*, 1379.

III. Ramos, J., Lopes, R. J., & Araújo, D. (2019). Soccer Data Revealing Both Design and Zipf-Mandelbrot Like Regulaties. Submited to *Complexity*.

IV. Ramos, J., Lopes, R. J., Marques, P., & Araújo, D. (2017). Hypernetworks: Capturing the Multilayers of Cooperative and Competitive Interactions in Soccer. Complex Systems in Sport, International Congress, 5-7 October, Barcelona. Linking Theory and Practice. Frontiers Abstract Book. ISBN: 978-2-88945-310-8 DOI: 10.3389/978-2-88945-310-8.

Other Publications (ISI indexed journals):

V. João Ribeiro, J.; Davids, K.; Araújo, D.; Silva, P.; Ramos, J.; Lopes, R.J.; Garganta, J. (2019). The Role of Hypernetworks as a Multilevel Methodology for Modelling and Understanding Dynamics of Team Sports Performance. *Sports Medicine*.

VI. Ramos, J., Lopes, R. J., & Araújo, D. (2019). Hypernetworks multilevel analysis of goal scoring opportunities in a soccer match. *In preparation.*

Other Publications (non-indexed journals):

VII. Araújo, D., Silva, P., & Ramos, J. (2014). Affordance-Based Decisions Guide Team Synergies During Match Performance. *Research in Physical Education, Sport & Health*, *3*(1).

Other Publications (book chapters):

VIII. Araújo, D.; Ramos, J.; Lopes, Rui J. (2016). Shared affordances guide interpersonal strategies in sport teams. In Interpersonal Coordination and Performance in Social Systems. *Pedro Passos, Keith Davids, Jia Yi Chow* (Ed.s). Routledge

Other Publications (conferences):

IX. Ramos, J., Araújo, D., & Lopes, R. J. (2012) Análise de Redes Sociais como Instrumento de estudo de desportos colectivos com bola: uma aplicação ao Voleibol. Simposium in *XIII Jornadas da Sociedade Portuguesa de Psicologia do Desporto*. Universidade Lusófona – Lisboa.

X. Ramos, J. (2014) Simposium organizor in *XV Jornadas da Sociedade Portuguesa de Psicologia do Desporto*: “Teoria das Redes, como Instrumento de Análise do Desempenho em Futebol – O caso do Mundial de Futebol 2014”.

XI. Ramos, J.P., Lopes, R. J., Araújo, D., Silva, J. P., & Silva, C. M. (2014). Equipas de Futebol como Redes Complexas – O caso da equipa campeã mundial. Simposium in *XV Jornadas da Sociedade Portuguesa de Psicologia do Desporto*. Escola Superior de Desporto de Rio Maior – Instituto Politécnico de Santarém. Rio Maior – Santarém.

XII. Lopes, R. J., Ramos, J. P., Araújo, D., Silva, J. P., Silva, C. M., & Serrão, C. (2014) Metodologia e ferramentas para recolha e processamento de dados em eventos desportivos - aplicação ao campeonato mundial de futebol. Simposium in *XV Jornadas da Sociedade Portuguesa de Psicologia do Desporto*. Escola Superior de Desporto de Rio Maior – Instituto Politécnico de Santarém. Rio Maior – Santarém.

XIII. Silva, J. P., Silva, C. M., Araújo, D., Ramos, J. P., & Lopes, R J. (2014) As redes sociais como ferramenta de análise do desempenho ao serviço do treinador. Simposium in *XV Jornadas da Sociedade Portuguesa de Psicologia do Desporto*. Escola Superior de Desporto de Rio Maior – Instituto Politécnico de Santarém. Rio Maior – Santarém.

XIV. Ramos, J., Araújo, D., & Lopes, R. J. (2016). Híperredes multinível revelam propriedades da dinâmica colectiva do jogo de futebol. Atas 1º Congresso Ibero-americano de Desporto, Atividade Física, Educação e Saúde.

XV. Ramos, J., Lopes, R., & Araújo, D. (2015) Análise de desempenho de equipas de futebol através de híperredes multinível: determinação da estrutura mínima representativa do jogo nos níveis micro-macro das equipas. Simposium in *XVI Jornadas da Sociedade Portuguesa de Psicologia do Desporto*. Instituto Politécnico da Guarda. Guarda.

XVI. Ramos, J., Lopes, R., & Araújo, D. (2017) A dinâmica das interações de cooperação e competição num jogo de futebol captadas por híperredes.

Simposium in *XVIII Jornadas da Sociedade Portuguesa de Psicologia do Desporto*. Instituto Politécnico de Bragança. Bragança.

Contents

List of Figures

1 General Introduction

The PhD realm is the study of sports performance analysis (PA) under the lens of complex systems studies. This PhD choice was due to difficulties of PA in sports using mainly Sports Sciences Approaches (notational analysis and motion analysis in Biomechanics focusing on individual performance). In this context what this thesis put forward is to tackle the large amount of data produced by the advances in technology using complex systems tools like: different levels of analysis (micro-meso-macro); hypernetworks (*n-ary* interactions between more than 2 players/actions) and multidisciplinary approaches (and team), allowing to overcome some limitations of more traditional approaches producing useful contributes to team sports researchers and practitioners.

The understanding of complex systems sciences it's a whole new paradigm that leads us to explore new frameworks and tools in a specific and relatively closed context like sports. Our voyage through this process began with the Doctoral Program Director advice to use the networks approach to study sports performance. Before having supervisors, I had produced an empirical study via the final report of the curricular year applying Social Network Analysis (SNA) and producing data from the Olympic volleyball women's final in London's Olympics Games and also volleyball matches from Physical Education level classes. In this first approach with SNA, I used NODE XL (plugin from Excel) and it was surprisingly promising allowing the production of volleyball networks with their corresponding metrics and present its results to some pairs (doctoral and master students in sports sciences at SpertLab in FMH-UL). In the next academic year I had presented an update of the study, with a comparison to previous Olympic women's volleyball final, in a workshop in the same context to some doctoral and masters students' pairs.

The next step was to invite two supervisors with different background and forming a multidisciplinary scientific team, combining networks and sports performance analysis. Our choice was in ISCTE-IUL and IT with Professor Rui Lopes (networks specialist) and FMH-UL with Professor Duarte Araújo (PA specialist). Our first task together was the thesis plan, accepted in 2013 march. This was the first complex (emergent) behavior of this multidisciplinary team, where each one's role began to be defined. From here, every paper or work produced involved the specific knowledge of all members of the team.

The next significant step was during Soccer World Championship 2014 in Rio-Brazil, where we had accepted an invitation of a daily national paper (Journal "O Público") to produce scientific commentary of the National Team matches based on a social network analysis, and published it in the next day paper and online editions. We had the additional participation of two master students: José Pedro Silva and Carlos Silva, to produce data from the events. Our method first steps consisted (see Figure 2.) on: i) with a specific software to register online (through television emission of each match) the interactions between players (ball flow passes) and players' actions (e.g. kick to goal, turn over); ii) the data produced was sent every 15 minutes to the production of output csv files; then we convert the csv files into xlsx files and use it on NODE XL (a plugin from EXCEL) and produce the networks graphs of each 15 minutes, half of the matches and the entire matches, and then commented through networks graphs; Afterword's a final version was produced and sent to Journal "Público" in order to be published (see figure 1.).


P2 ÍPSILON CULTO FUGAS P3 CINECARTAZ

DESPORTO › SL BENFICA FC PORTO SPORTING FUTEBOL NACIONAL FUTEBOL INTERNACIONAL RÂGUEBI TÉNIS MOTORES GOLFE RUNNING OUTRAS MODALIDADES

CRISTIANO RONALDO

EUA-Portugal à lupa: pouco ataque pelo centro e pouca precisão no remate

Equipa académica analisou a forma como os jogadores interagiram no jogo que quase ditou a eliminacão da selecção portuguesa do Mundial 2014.

23 de Junho de 2014, 16:18

Projecto do Laboratório de Perícia no Desporto da Faculdade de Motricidade Humana da Universidade de Lisboa em parceria com o Programa Doutoral em Ciências da Complexidade (ISCTE-IUL e FCUL).

Coordenador
Duarte Araújo (FMH-UL)

Equipa
Rui Lopes (ISCTE-IUL e IT-IUL)
João Paulo Ramos (ISCTE-IUL e FEFD-ULHT)
José Pedro Silva (FMH-UL)
Carlos Manuel Silva (FMH-UL)

Figure 1. EUA-Portugal à lupa: pouco ataque pelo centro e pouca precisão no remate. Lisbon: Jornal Público Online. 2014 (Araújo, Lopes, Ramos, Silva, & Silva, 2014). https://www.publico.pt/2014/06/23/desporto/noticia/euaportugal-a-lupa-pouco-ataque-pelo-meio-e-pouca-precisao-no-remate-1660182. Accessed 21 Jun 2019.

When Portugal was eliminated in the group phase, the Journal invited us to follow some major candidates to winning the world championship. The method and results were later presented in scientific meetings and workshops on performance analysis with networks approach (see publications/conferences items X, XI, XII and XIII).

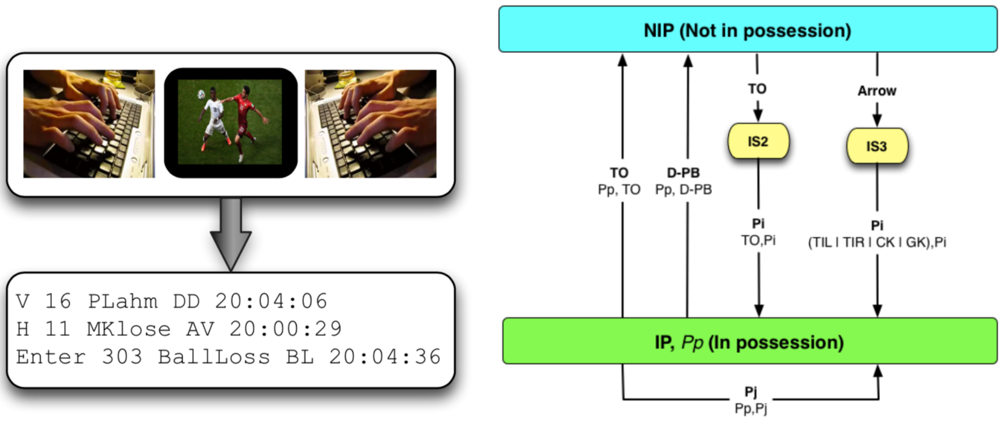


Figure 2. Collecting data in a *state machine* (one for each team): two "steady" states (NIP or IP) and intermediate states (registering events [TO – turn over and D-PB – loss of the ball] and players [Pi]) that lead to state changes (actions like ball pass [Pp, Pi] or shot to goal).

The experience of presenting the method and results to coaches and investigators and listening to their criticism about the limited interest of the results have nudged us to search for a more holistic and dynamical approach to sports teams as complex systems. At this time, we began to produce a state-of-the-art paper on networks in PA and published it in Sports Medicine (Ramos, Lopes, & Araújo, 2017) (found in Section 3.1). In this paper, we have reviewed social network analysis (SNA) in the PA context and exposed some common pitfalls related to the direct and not sufficiently grounded use of some SNA metrics in the case of team sports competitions. We have also discussed that this approach to PA was focusing on the *dynamics on the networks* and not on the more interesting *dynamics of the networks*. In this paper is putted forward the proposal that an appropriate way of addresses this latter issue can be achieved through complex dynamic networks concepts, notably: bipartite (multilayer) and temporal networks.

Consequently we found a possible answer to some of SNA's shortcomings by using multilevel hypernetworks.

The first known study in sport contexts introducing structural analysis with more than two interactions (based on Atkin, 1972)) and considering competition interactions, was only in 1980 (Gould & Gatrell, 1980). This paper from Gould and Gatrell (1980) however, had small relevance due to the reduced number of citations. With a gap of almost 30 years, we could see some applications of hypernetworks in PA in robotic soccer matches by Jeffrey Johnson's (Johnson & Iravani, 2007) . In the current thesis work, multilevel hypernetworks approach (MHA) to soccer matches is based on real world soccer players' positional data from eleven matches of the English Premier League in the season of 2010-2011, obtained from the former Prozone (currently STATS). We produced and published a paper (Ramos, Lopes, Marques, & Araújo, 2017a) describing this work in Frontiers in Psychology (open access) special number (found in Section 3.2). This paper focus is different from Johnson's works, given we had: i) established a non-parametric criteria for forming the sets of players based on their interpersonal distance (closest player); ii) identified and analyzed the most frequent simplices and where they occurred (in the match field); iii) identified local dominance (in terms of the numbers of players from each team) and the players' moves to achieve this dominance, and iv) a multilevel analysis of the dynamics of the simplices of simplices formed in some specific events. Our next studies with MHA evolved through a more detailed analysis of the structures and dynamics of each level of analysis, which allowed the identification of the simplices and the players' that constitute them. The results from these advances were presented at the International Congress of Complex Systems in Sports (Barcelona) and the resulting extended abstract was published in the book of abstracts by Frontiers (Ramos, Lopes, Marques, & Araújo, 2017b) (see Section 3.3). The mathematical formalisms of the simplices was also matured, allowing the representation of the structure and dynamics of the simplices, introducing the ball carrier representation and the relative position of the players' regarding the goal. This step was also presented in a national scientific symposium (publications/conferences - XVI).

It is worth mentioning that the regular participation in this type of national symposiums where this approach was presented led to a significant interest on it by the community. Notably, the work developed by Ribeiro and colleagues (Ribeiro et al.,

2019) overcame one of our data limitations (ball positioning) and is an important contribution to our line of investigation regarding MHA. In this co-authored paper we could propose some advances of MHA, specifically our contribution to the paper was in the hypernetworks mathematical formalisms that included more detailed information of player actions on disaggregation and aggregation of simplices and regarding the ball possession.

In the final stage of this PhD journey, we observed very interesting statistical results when analyzing the simplices' set occurrence distribution. These distributions obtained from ten soccer matches revealed, on one hand, well known models and empirical laws, such as the Zipf-Mandelbrot law (ZM), and on the other hand, the possible impact of design, i.e. match strategy, on the simplices' set statistical distribution. Soccer teams' match strategy is designed like all complex social systems through purpose (e.g. the simplices goal-keeper and goal) and intentionality (e.g. the simplices left defender and right attacker from opposite teams). The analysis of the goodness of fit of the ZM model through the chi square validity test revealed that in all ten matches' analyzed there were "exceptions", i.e. outliers to the model. These were found typically in the two to five most frequent simplices and after looking at their composition this appeared as an expression of complex social systems' design. A paper was submitted to Scientific Reports – Nature presenting these findings, that are at the core of the realm and contribution of this thesis.

Finally, we have prepared a paper were those last achievements in simplices formalisms are applied to critical events like goal scoring opportunities (GSO). In this paper we explore the dynamics of the simplices forming and disaggregation, some seconds before GSO. We use inertia concepts to explain how the players' moves are producing the expansion or contraction of the simplice's area and therefore contributing to the disaggregation or sustentation of the simplex.

1.1 Complexity Sciences in team sports performance

Complexity sciences' approach to team sports performance via complexity sciences tools like network theory implies that team sports have systemic properties and therefore exhibits complexity (Juarrero, 2010).

A common criteria for identifying a system as complex is that it must have their parts interconnected or interwoven (Bar-Yam, 1997). Therefore, complex physical or social structure will not emerge in a weakly interdependent or even independent parts (Juarrero, 2010). In this way, one system can increase their complexity by either increasing the number of their parts (structure) and, or increasing their interactions (functionally) (Bar-Yam, 2004). Thus, this feature is associated to unexpected or unpredictable emergence of the system (Johnson, 2013; Johnson & Iravani, 2007). Usually in sports settings the number of parts are equivalent and predefined, therefore, a greater complexity depends on the interactions between sub-systems (teams) members, which in sports context results in more unpredictable teams (Bar-Yam, 2003).

One the other hand another significant feature is that, most of the behaviors of complex systems emerge from the adaptability to the changing environments/constraints (e.g., players fatigue, position in the field, score changes, players substitutions) (Balague, Torrents, Hristovski, Davids, & Araújo, 2013; Johnson & Iravani, 2007). This adaptive property to the continuously changing constraints promotes the emergence of new forms of behavior (creative or innovative) that were not imposed or previously designed (Balague et al., 2013). According to Bar-Yam (2003), when one player or team has a more diverse set of offensive plays (s)he, or team becomes more unpredictable for the defenders, which results in additional loss of energy for those.

In sports settings the observed complex systems exhibits structurally and functionally heterogeneous components (players) that interacts in different spatio-temporal scales and with varying intensities (Balague et al., 2013). Scalability is referred as a central element of complexity science, once many complex systems are organized on multiple levels and exhibiting the same dynamics across levels (Juarrero, 2010). This means that with the right tools we can access the current state of the system, e.g. at the macro-level (the level of the whole system) and understand its dynamics produced by the interaction of the elements at lower levels of the system (Johnson, Fortune, & Bromley, 2017). Thus, we can understand that complex systems have many sub-systems that can be described in several levels of organization, varying from physical and social multilevel subsystems, with their own intra-level and inter-level bottom-up and top-down dynamics (Johnson, 2013). For Johnson (2013) it still lacks a formalism that represents multilevel dynamics (Systems of Systems of Systems) and it remains an obstacle to scientific progress.

The analysis of the different levels in social processes like team sports competitions is typically that of: the microlevel corresponds to individuals (e.g. teams players); the mesolevel represents the structures (e.g. different teams and, or sets of players); and the macrolevel represents the dynamics of the whole system (e.g. specific events like goal scoring opportunities) (Johnson et al., 2017; Ramos, Lopes, et al., 2017a).

The correlations between different levels of scale are one of the reasons behind the order that emerges in complex systems, and, at each level, the organization of phenomena has not a preferred scale or dimension. Therefore, the morphogenesis that evolves through levels of scale can not be explained by the exponential nor the Normal distribution (Komulainen, 2004; Salingaros & West, 1999). When considering these scaling phenomena, one way of begin to understand it is to apply "The Least Effort Hypothesis" that described the minimization of the efforts of speaker and hearer proposed by Zipf (Piantadosi, 2014; Zipf, 1949). This has a result the of speaker's tendency to use a few words; and the tendency of hearer to demand a specific word. This hypothesis is the basis for one type of power scaling distributions in complex systems. This type of power law is related to discrete distributions and describes well different phenomena like frequency of words, firm sizes, city sizes (Piantadosi, 2014), and the goals scored by players in different championships (Malacarne & Mendes, 2000). When applying it to sports like soccer, we pose the research question if cooperation and competition interactions would promote such a distributions, in the sets of players (proximity positioning) in the entire matches.

One complex social systems feature that is not the result of emergence is the influence of design. For Johnson (2013), when the systems are created by human's, they reveal some artificiality due to the need of achieving some specific outcomes (Alexiou, Johnson, & Zamenopoulos, 2009; Johnson, 2013). This artificiality is related with the idea that not everything is emergent and adaptive, but in some part also predefined and designed to happen. Sports coaches' need to explore all those constraints in order to maximize the productivity of their athletes and teams. This means that there are strategies (predefined purpose and intentions on cooperative and competitive interactions) that constrains players' moves during the matches. For a better understanding, besides goalkeepers "attraction" to their goals (considering goals also one interacting part of the system), we could find some sets of players (typically one defense and the closer attacker) that are closer to each other for a large amount of time or moments.

In summary, the approach of complex systems to team sports performance identifies several features like: the systems' many heterogeneous parts; its dynamics emerging from interactions of autonomous agents; the unexpected or unpredictable nature of emergence; multiple subsystem dependencies; self-organization into new structures and behaviors; adaptation to changing environments; co-evolving subsystems; multilevel dynamics; statistical systems regularities (ZM like); unrepeatable experiments and design (Balague et al., 2013; Johnson, 2013; Juarrero, 2010).

The continuous interest from coaches and athletes in better understanding the dynamics of the team performance during the competition and implementing training plans that increases team performance, has motivated investigators from all areas to pursue to find better tools and theoretical frameworks that better explains collective behavior in team sports as complex systems (Balague et al., 2013; Zhu et al., 2009).

The systematic research on those theoretical frameworks and tools has been influenced from complexity sciences and evolved from three main approaches: coordination dynamics (which studies how changing constraints influences behavioral pattern formation and what principles and laws explains it); ecological dynamics (where the relevant level of analysis and explanation is the performer-environment system) (Duarte Araújo & Bourbousson, 2016); and, network theory and tools (networks of players interacting in a cooperative and competitive way, exchanging tokens - typically a ball, and moving in a limited space and time (Balague et al., 2013; Ramos, Lopes, & Araújo, 2017). The later overtake the heavy jargon and complex research methods and emerged as a useful and appropriate tool due to its visual communication power (Balague et al., 2013; Fewell, Armbruster, Ingraham, Petersen, & Waters, 2012; Glöckner, Heinen, Johnson, & Raab, 2012; Johnson & Iravani, 2007; Juarrero, 2010).

In the next chapter we analyze how complex systems approaches influenced PA in sports.

1.2 Complexity Approach to Performance analysis

Performance analysis has emerged in the last decade as a sub-discipline of sport sciences. Despite its success, that began with biomechanics and notational analysis in describing the performance trends of players and teams notably their strengths and weaknesses (Vilar, Araújo, Davids, & Button, 2012), it has been pointed out that there is some reductionism on their functional utility in some specific performance situations (in several sports). These are pointed out by omitting references to “who” and “why” (Vilar et al., 2012), and also for considering single observations independently of the previous ones, which implies considering teams and individuals in isolation (Bartlett, Button, Robins, Dutt-Mazumder, & Kennedy, 2012; McGarry, 2009), or the limitation about the understanding and enhancement on application by the coaches and athletes (Glazier, 2010).

Considering these limitations and criticisms, one of the main challenges for performance analysts it is on identifying the possible common patterns in games that best explain the overall performance of teams (Dutt-Mazumder, Button, Robins, & Bartlett, 2011; Glazier, 2010). It has been pointed out that only by considering them as complex non-linear dynamical systems it is possible to move towards examining player and team interactions (Bartlett et al., 2012; McGarry, 2009). The coordination dynamics between players and teams in competition is usually analyzed between two entities (e.g. two players or two teams coordination patterns):

i whether the dyad is between opposing team members (attacker-defender) trying to maintain (defender) their symmetry;

ii. or to break it (attacker) in order to accomplish their goal, like promoting free-up space if in attacking phase, or tie-up space if in defensive sub-phase (Araujo, Davids, Bennett, Button, & Chapman, 2004; McGarry, Anderson, Wallace, Hughes, & Franks, 2002; Reilly, Cabri, & Araújo, 2005; Williams & Hodges, 2004). This search for coordination is also found between team members, whether in offensive or defensive phase, increasing complexity due to the addition of further intentional and informational constraints, in mutually related spatial and temporal aspects of interacting movements. However this approach is still failing in providing enough and accurate information about the performance context, helping practitioners in their intervention (Araujo, Correia, & Davids, 2012). According to Hughes and Franks (2008) the tactical evaluation of the athletes and teams performance has dependency on their opponents, meaning either the dyadic attacker-defender (Araujo et al., 2004; McGarry et al., 2002; Reilly et al., 2005; Williams & Hodges, 2004) or team attacking vs. team defending collective behaviors. One of the approaches that has been proposed for studying these relations is Social Network Analysis (SNA) (Clemente, Couceiro, Martins, & Mendes, 2014, 2015; Clemente, Martins, Kalamaras, Wong, & Mendes, 2015; Duarte et al., 2012; Duch, Waitzman, & Amaral, 2010; Dutt-Mazumder et al., 2011; Grund, 2012; Passos et al., 2011; Ramos, Lopes, & Araújo, 2017). Its potential lies not only in its visualization properties but mainly because it analyzed the dyads, i.e. pairs of actors rather than a monad (a singleton actor). This is in concordance with several investigators (Brandes, Freeman, & Wagner, 2012; Grund, 2012) when they argue that the focus of team performance must be not on individual team members but rather on the dyads between team members (the orchestration of interactions and their interactions). Some of the studies using this approach have already made some advances through qualitative and quantitative study of the relation between the structural properties, like network centrality (high centrality values associated to less efficient team performances) and density (high intensity in network associated to better team performances) of interactions between team members and the performance outcomes (Balkundi & Kilduff, 2006; Grund, 2012; Katz, Lazer, Arrow, & Contractor, 2004).

However, in what way may in Social Networks perspective lead to greater performance, especially in team processes, is still unclear (Duch et al., 2010), remaining the uncertainty on how individual strengths and roles (including "superstars") are combined for optimal results. Typically, this approach considers static network structure that is obtained via the aggregation of all the interactions (passes) that occurred in an entire match (Ramos, Lopes, & Araújo, 2017). This may lead not only to concealment of important concepts such as attacking play but also to metrics that may be misleading (Ramos, Lopes, & Araújo, 2017). A notable example is metrics based on `shortest paths´ over the aggregated network, that occur in two scenarios: i) pass interactions form a walk and not necessarily a path; ii) interactions between players in a match do not follow this principle but may be better described by geographic networks and random walks (Ramos, Lopes, & Araújo, 2017).

In our review (state-of-the-art) published paper (Ramos, Lopes, & Araújo, 2017) (found in section 3.1.), we identify and tackle some of the pointed limitations from the use of Social Networks Analysis. At the first, we point out that most of the approaches are based on the *dynamics on the network*, which focuses on flows across the network structure, e.g. ball passes between players. On the other hand, the *dynamics of the network* are concerned with changes in the network structure itself, e.g. how the players' positions can provide information on a possible pass or player action. In the former case, the processes within the network are lost by representing only the "ball flux" as static *flip books*, where node position remains constant but edges cumulate over time, and therefore the researcher or practitioner obtains only cumulative snapshots of the network as a function of time (Moody, McFarland, & Bender-Demoll, 2005; Ramos, Lopes, & Araújo, 2017).

We have also referred that the dynamic relations in team sports games have other basic dimensions, in addition to passes, that have not yet been fully captured in sports settings, notably: i) relational space (i.e. interactions considered in a geographical space); ii) their time structure (i.e., rate change, order or sequence, or simultaneity of interactions); and iii) their relations with the different types of nodes involved (i.e. colleagues or adversaries), thus considering both cooperation and competition interactions (Moody et al., 2005; Ramos, Lopes, & Araújo, 2017). Thus, network dynamics analyses may explain how and why disequilibrium situations such as scoring opportunities occur. PA experts point out this and other research questions that remain unclear, like exploring ways of exhibiting the dynamic interrelations between tactics adopted by the different types of athletes and teams within the same competition (game sub-phases) and through the diversity of outcomes (Hughes & Franks, 2008).

In that paper we put forward the proposal of using bipartite networks to identify nodes representing players and technical actions in different layers that can be extended hierarchically. Technical actions (level *n* events) can be linked in another bipartite network to level *n+1* events corresponding to higher-level concepts. This multilayer approach can be extended to all other relevant types of interactions. Either for cooperative (e.g., between players of the same team in order to create a scoring opportunity) or competitive interactions (e.g., between players of different teams competing for ball possession) and that may be captured and analyzed via multilevel hypernetworks. In addiction and supporting our proposal, according to Boccaletti and colleagues (2014), multilevel networks represented a major advance in many areas of science since as they describe systems that are interconnected through different categories of connections (e.g. relationship: teammate vs. opponent; activity: increasing vs. diminishing interpersonal distance; category: attacker vs. midfielder) and therefore can be represented in multiple layers, including networks of networks (e.g. interactions between teams). A third contribution from the paper is the use of hypernetwork for representing interactions and relations in the sport match, as, a hyperedge can connect more than two nodes, thus directly representing *n-ary* interactions occurring among small sets of nodes, $\langle p_i, \ldots, p_j \rangle$ (Boccaletti et al., 2014; Criado, Romance, & Vela-Pérez, 2010; Johnson, 2006, 2008, 2013, 2016; Ramos, Lopes, et al., 2017a). This generalization provided by hypernetworks enables the representation of cooperative and competitive interactions that occur during the game and that involve an arbitrary number of players (either teammates or opponents).

A seminal approach to the use of complex systems tools in soccer analysis via multi-dimensional analysis was published by Gould and Gatrell (Gould & Gatrell, 1980). Here they have used a structure analysis based on polyhedral dynamics (Atkin, 1972; Atkin, 1974; Atkin, Hartston, & Witten, 1976; Atkin & Witten, 1975) and considering the cooperation interactions through passing the ball and the competition interactions regarding the stealing of the ball. This paper introduced several ideas that we have developed and extended in our work. Notably: defining *n-ary* sets of players, connected through passing or stealing interactions; having goals as pseudo-players; and finally considering the more stable structure (backcloth) and the more dynamical interactions inside those structures (traffic).

The interest in PA has also expanded to Artificial Intelligence (AI), namely in Robot soccer which confirms the relevance of PA as a scientific area. AI scientists involved in robot soccer have set a long-term goal in order to promote investigation in this field, and that is that: by the year of 2050, a humanoid robotics team should be capable, according to FIFA rules, to defeat the world champion human team in a soccer match (Kitano, Asada, Kuniyoshi, Noda, & Osawa, 1997). We can identify some common subjects (related to investigations in human soccer) that are addressed in the existing literature of AI regarding the investigation of PA in robotic soccer, notably: goal-scoring behaviors through cooperation interactions (either pre-designed or adapted in robotic soccer) (Almeida, Abreu, Lau, & Reis, 2012); comparing robotic soccer matches with human ones (actions, tactics, statistics) (Abreu et al., 2012); develop a more precise and controlled kicking (Barrett, Genter, Hester, Quinlan, & Stone, 2010); motion analysis (Abreu, Moura, Silva, Reis, & Garganta, 2012); and, the game strategy description and prediction of ball motion (Martinovic et al., 2010). The relevance of these studies consists on the efforts made by AI in incorporating various technologies and achieving innumerous technical breakthroughs (Kitano et al., 1997). Some are relevant for human soccer PA, namely those that allows the development of automatic PA systems. By incorporating the big data provided by the cartesian coordinates of players' and ball, the related motion analysis systems require complex algorithms (e.g. multiagents collaboration, strategy acquisition, real-time reasoning and planning (Kitano et al., 1997)) that are being developed and tested by researchers of AI in simple simulated environments like "RoboCup" (Abreu et al., 2012; Martinovic et al., 2010).

The idea of artificiality is present in complex social systems, like soccer matches, when the team coaches design the strategy for the match, constraining the intentions and decisions of the players for the match. This idea adjusts perfectly in the studies of PA using AI in robotic soccer, considering that the human design is also present in the two phases of preparation for a robot soccer match. The offline phase, where the main goal is to detect opponent play patterns (in previous matches) and pre-define the strategy to neutralize them. The online phase is to adapt the team strategy regarding the analysis of the opponent behavior during the match (Almeida et al., 2012). However, there are still significant differences between human soccer and robotic soccer (Abreu et al., 2012).

Therefore, our research questions for this thesis are:

What are the main factors that influence sports team effectiveness and performance?

What are the structural and dynamical properties in cooperative (synergetic) and competitive interactions that most influence their performance outcomes?

Is complex networks approach and its related tools able to identify, in different levels of analysis, the structure and the dynamics of the cooperative and competitive interactions in team sport complex systems, considering the results, the classifications and the match time?

Are the results obtained from complex network analysis useful in training/preparing situations? What are the structural properties that can help transmitting ecological validity from training to competition?

What are the relevant types of interactions for the analysis of the structure and dynamics of the cooperative and competitive interactions between team players?

1.3 Thesis structure

This document follows the "three papers" model and in its structured in four main chapters. In chapter one – Introduction, we put forward the thesis and provide a general introduction, the theoretical mainframe and the empirical context of the thesis realm. Chapter two – State-of-the-art, is dedicated to exploring the literature on the multidisciplinary approach used. Namely, it describes complexity sciences studies usage of complex networks theory and performance analysis in sports. In the third chapter are presented the four papers that correspond the core of the thesis. For each paper is described it context and summary, followed by the paper itself (in a *verbatim* copy). There is, one review paper on network theory in sports; one paper focusing on using Multilevel Hypernetworks Theory (MHT); one conference paper/extended abstract on empirical results from MHT and its implications on practice; and finally, one paper where is described how soccer matches present complex systems features like:

i. Zipf-Mandelbrot regularities of the empirical data, on cooperation and competition interactions within soccer opposing teams;
ii. Design as a complex social systems emergent feature in some particular sets of players formed through spatial positioning.

The last chapter discusses what are these thesis' work main contributions, practical implications and applications, limitations and suggestions for future works.

2 State-of-the-art on Performance Analysis, Complex Networks and Complexity

Contemporary organizations adopted teams as their basic unit of work (Balkundi & Kilduff, 2006), which explains why teams have such prevalence on organizations. Therefore, team performance became so important that for almost the last three decades had evolved and matured as a science. The knowledge produced by this science, has implications on managing teams, groups, crews and collectives, and given evidence-based principles, guidelines, tools, methodologies and specifications applied to many domains (Salas et al., 2010).

2.1 Performance Analysis in Team Sports

Team sports are one clear example where participation cannot be separated from the intention to improve performance, and that's one domain where managing teams or groups has been largely studied (Hughes & Franks, 2008).

In sport sciences the improvement on performance studies focused mainly on how coaches and athletes could better analyze the performance (Hughes & Franks, 2008) and influence a better feedback on the knowledge of the results (KR) or performance (KP). This resulted in the emergence of an independent sub-discipline of sport science, known as *performance analysis* (Glazier, 2010), which main goal is to provide accurate and augmented information to coaches and athletes for the future improvement of team performance (Vilar et al., 2012).

The management of team performance has evolved from inaccurate and unreliable subjective observations from coaches, due to limitations of memory, problems of highlighting, and other observational difficulties. Despite these limitations, feedback (FB) and KR provide relevant benefits, notably the evolution obtained by providing FB with the use of video analysis and other computer-aided technologies resulted specially in high performance athletes. On the other hand, the need to define and identify the critical elements of performance to other levels of athletes became difficult when the challenges due to the complexity of the systems to be observed were such that the real-time notation was not practically feasible (only with slow motion and replay was possible to overcome some of those difficulties). Since the early stages, factors for performance of sport teams have include the management of information complexity, addressing the reliability and validity of data and exploiting artificial intelligence approaches and methods for its processing (Hughes & Franks, 2008; Hughes & Bartlett, 2002).

2.1.1 Traditional sub-disciplines in performance analysis

Investigators on biomechanical and notational analysis[1] became the main producers of investigation in this area, developing theoretical models based on performance indicators, with the extensive use of video analysis and technology, creating systematic techniques of observation for coaches and athletes to benefit of a more precise feedback (Hughes & Franks, 2008).

Sport biomechanists have concentrated their studies on sports that involve mostly closed skills because of the importance of movement technique, like in acrobatic, athletic and cyclic sports. Typically, the performance parameters are angles of attack and release of critical movement techniques (Hughes and Bartlett, 2002). Therefore this analysis has focused mainly on isolated individual closed skills even in the few examples where it has been applied to team sports such as cricket (Bartlett, Stockill, Elliott, & Burnett, 1996), in soccer (Lees & Nolan, 1998) and in rugby or racquet sports (Hughes & Bartlett, 2002).

On the other hand, notational analysts had been focusing on general match indicators, tactical indicators and technical indicators that are more related to interactions between players and the movements and behaviors of individual team members (Hughes & Bartlett, 2002). The performance indicators parameters for sport biomechanists, are in notational analysis used to access the performance of an individual, a team or elements of a team (Hughes & Franks, 2008). The interest in the methods of notational analysis, both from practitioners and investigators, is due to its successful description of performance tendencies of players and teams, namely their strengths and weaknesses in some specific situations, using scoring indicators (e.g. goals, baskets, winning shots, errors, ratios of winners to errors and goals to shots); and performance indicators like action frequencies of players and teams (e.g. turnovers,

[1] "Notational analysis is a method of recording and analyzing dynamic and complex situations such as field games" (Hughes & Franks, 2008, pp.181).

tackles, passes/ball possession, etc.) that are usually associated to successful performance (Vilar et al., 2012).

Although these methods provided useful information for the practitioners in the form of performance indicators (or parameters) some limitations became emerged regarding their functional utility (Duarte et al., 2012; Glazier, 2010; Vilar et al., 2012), also some criticisms once it is not clear how much these variables significantly enhance team performance or benefits coaches, athletes and teams (Glazier, 2010). In the same lines Hughes and Franks (2008) had previously pointed out the need for the tactical evaluation of the athlete performance to also represent their dependency on opponents; namely, that research should explore ways of exhibiting the dynamic interrelations between tactics adopted by the different types of athletes within the same competition.

The generally accepted conceptualization that is common to sport biomechanics and notational analysis is that they are measuring and describing the same emergent pattern formation but at different scales of analysis, resulting in a fragmented application and in a lack of explanatory power (Dutt-Mazumder et al., 2011; Glazier, 2010). The reasons for this fragmentation and consequently their descriptive rather than explanatory outcomes, is on the lack of a theoretical framework that can integrate them hopefully with others sub-disciplines of sport science that are concerned with enhancing performance, like physiology, psychology and motor control (Dutt-Mazumder et al., 2011; Glazier, 2010). Another aspect that should be considered is that research on performance analysis is far from the accurate prediction of human motor performance in a given task at a given time, due to the complex, non-linear, interactions between, not only the many independent component parts of the human movement system but also, all the different levels of the system considering also their surrounding environment and the specificity of the tasks undertaken (Glazier, 2010).

2.1.2 Dynamical Systems Theory in Performance Analysis

According Glazier (2010) and Dutt-Mazumder and colleagues (2011) the unifying theoretical framework that has the power to integrate the referred sub-disciplines of sport science is dynamical systems theory (DST). However the technical jargon of this approach can be discouraging for practitioners starting with the many *degrees of freedom* which are the many independent parts that are free to vary over space and time; continuing with complex systems that are typically *open systems* operating under conditions far from equilibrium, meaning they interact with the environment and are in a constant state of flux due to changes in internal and external energy flows; and, having also an enormous potential for disorder which implies to be able to exploit these energy flows and the surrounding constraints to form orderly and stable relationships among the many degrees of freedom at *different levels of the system*. Other concepts are for instance, the idea that *attractor* states, meaning functional coordinative states *emerging spontaneously* in a physical *self-organizing* processes rather than being pre-planned (top-down) by an executive intelligent that controls everything, and once assembled (this functional coordinative states) the many independent parts operate autonomously and searching for self-regulation provoked by internal or external perturbation, in order to preserve the system output and therefore being functionally coupled with the task (Glazier, 2010).

Synergetic Strategy

For (Glazier, 2010) the difficulties concerning the technical jargon and its impairment to an effective contribution to practical contribution has been mainly referred in the use of the synergetic strategy. This is a research strategy commonly used by human movement scientists to study pattern formation in complex neurobiological systems. The initial works on synergetic in neurobiological systems by (Haken, 1983) were in search of what could move the system through its many different coordinative states. That is, what order parameter (Kelso, 1995) or collective variables define stable and reproducible relationships among different degrees of freedom and control parameters (Kelso, 1995). Therefore, in this type of systems, *relative phase* has been the (only) order parameter (Michaels & Beek, 1995) and *oscillatory frequency* the important control parameter (Haken, Kelso, & Bunz, 1985; Kelso, 1984; Kelso, 1995). Consequently, the main goal of synergetic strategy is the identification of control parameters, manipulate them, and observe the changes that are produced in order parameters and other non-linear phenomena (Glazier, 2010). The empirical analysis of intra-individual coordination (Kelso & Jeka, 1992; Kelso, Buchanan, & Wallace, 1991) and inter-individual coordination (Schmidt, Carello, & Turvey, 1990; Schmidt, O'Brien, & Sysko, 1999) are the most successful applications of this strategy (Glazier, 2010), although initially outside sports contexts. The synergetic approach provided relevant contributions in sports, published by Riley and colleagues (Riley, Richardson, Shockley, & Ramenzoni, 2011) and also from Araújo and Davids (Araújo & Davids, 2016) where teams and athletes are seen as co-evolving subsystems that self-organize into new structures and behaviors, i.e., they form team synergies (Ramos, Lopes, et al., 2017a). For Araújo and Davids (2016) when the degrees of freedom of the different individuals co-regulate each other in order to complete a specific task, we can observe the emergence of collective synergetic behaviors. When one player in a sports team influences other team members behaviors this is one important team synergy feature (Araújo & Davids, 2016). The identification and measurement of synergies in a sports team can be done through some key system properties, such as: dimensional compression; reciprocal compensation; interpersonal linkages and degeneracy (Araújo & Davids, 2016). Dimensional compression refers to the ability of reducing degrees of freedom in the entire team system synergy due to the coupling of the independent degrees of freedom (Araújo & Davids, 2016). Reciprocal compensation is observed when one of the team members decreases is productivity (e.g. due to fatigue or an injury) and other team members adjust their contributions in order to achieve the teams' common task (Araújo & Davids, 2016). Interpersonal linkages refer to the individual player contribution to a team task, namely through its unique and specific characteristics (Araújo & Davids, 2016). Regarding degeneracy in the sports context, one player can use different motor behavior without compromising the whole team function, which is also observed from the team perspective when players adaptively interact continuously to accomplish a shared goal (Araújo & Davids, 2016). Considering that mainly cooperative interactions are considered in synergies, hypernetworks could contribute as a valuable tool to analyze these collective behaviors based on *n*-ary relations between team members and adding the competitive interactions too (Araújo & Davids, 2016).

Ecological Dynamics

In the sport and human movement related literature, there are references to what is called the 'constraints based' approach (Araújo, Davids, & Hristovski, 2006; Davids, 2008; Davids & Araújo, 2010) proposing it as an alternative approach in performance analysis. According to this approach, the explanation on pattern formation in neurobiological systems emerges from physical and informational constraints that coalesce to shape coordinative states on the system, and influence competing and cooperating interactions (Glazier, 2010). Despite initial focus were, in single-agent neurobiological systems helping to explain emergent pattern formation, meaning intra-personal coordination; it has been helping to clarify the emergent pattern formation in multi-agent neurobiological systems, meaning inter-personal coordination (Glazier, 2010).

Extending this approach, several investigators (Araújo et al., 2006; Davids & Araújo, 2010; Vilar et al., 2012) explored a combination of DST with ideas from ecological psychology in order to evolve from the constraints based approach, which is clearly explained by (Kugler & Turvey, 1987) in their book: “Ecological Science, in its broadest sense, is a multidisciplinary approach to the study of living systems, their environments and the reciprocity that as evolved between the two” and therefore Ecological Psychology is “…the study of information transactions between living systems and their environments, especially as they pertain to perceiving situations of significance to planning and executing of purposes activated in an environment.". This way Vilar and colleagues (2012) were trying to better understand how the adaptive behaviors of players are constrained by the information available in the performance environment, and therefore have proposed ecological dynamics for the study of team games. This approach recognizes the inherent adaptive flexibility in achieving successful performance outcomes at two levels of athletes (degeneracy of neurobiological systems) and of sports teams (social neurobiological systems); thus explaining how from different movements (motor equivalence) or tactical patterns may emerge the same successful performance outcomes. Ecological dynamics had already provided powerful theoretical explanation of behaviors in those complex neurobiological systems (Araújo et al., 2006; Davids & Araújo, 2010), mainly on performer-environment relations. Mainly, through the functional patterns of coordinated behavior that emerge from the process of self-organization of those performers interactions with each other under the specific constraints, whether they’re task or environmental constraints. The coordination between performers, expressed through the interactions between players and the information provided by the performance environment constrains the emergence of stable patterns; or their variability (expressed by the loss of coordination); and may even lead to emergence of new patterns of coordination that result from symmetry-breaking in organizational states. Emergence and variability of coordination patterns are chief concerns for investigators of sport sciences and coaches in team performance analyses (Vilar et al., 2012). These studies of team performance have analyzed mainly the emergent patterns of coordination in attacker-defender subsystems, like one (attacker) vs. one (defender) or two vs. one situations, in order to capture system organization and its changes over time through the identification of collective variables (Vilar et al., 2012). The already referred *relative phase* (Michaels & Beek, 1995) has confirmed its potential for a collective variable as shown in a study from (Bourbousson, Seve, & McGarry, 2010) in basketball, where the dynamics of relative phase between dyadic system performers allowed the quantitative expression of coordination processes.

When one considers the relative positioning of an attacker with the ball and a marking defender near the goal area, basket or final line, a very common one versus one sub-phase in invasive team ball games, it and can be referred to as and studied as a dyad (Duarte et al., 2004; McGarry et al., 2002). The dyad formed by attacker and defender, plus the proximity to offensive goal, comprises a system. The aim of the attacker is to perturb and ultimately to 'brake' the stability of this system. The defender tries to matches the movements of his opponent, keeping in position between the attacker and the goal, in order to maintain stability in the symmetry of the system (Duarte et al., 2004).

Ecological dynamics is therefore concerned with the influence of the spatial properties in field games, such as the proximity to the finalization targets or to opponents and how these might constrain the coordination in dyadic systems (Correia, Araujo, Craig, & Passos, 2011; Correia et al., 2012; Vilar et al., 2012). These types of studies are typically conducted with the use of video analysis and technology applied to performance analysis.

The use of positional data used by coordination dynamics studies highlighted two another candidates of collective variables (Bartlett et al., 2012): *team centroid* and *stretch index* (Frencken, Lemmink, Delleman, & Visscher, 2011). The former represents the team's position, calculated from the mean of the players positions (Frencken et al., 2011); and the later is a measure for the dispersion or spread of the team, calculated using the average distance of the players to the team centroid (Bourbousson et al., 2010). According to Bartlett and colleagues (2012) these measures did not reveal to be sensitive enough to be associated to critical events (scorings or turn-overs), pointing out alternatively some ideas on multi-dimensional coordination for the use of networks (artificial neural networks, more specifically self-organizing maps).

Other studies have revealed how the interaction occurs during the entire competition, reinforcing the need for renovation from notational analysis methods (Duarte et al., 2012). Moreover, methods based on DST, like ecological dynamics showed that the coordination dynamics expressed by movement patterns in team sports present nonlinear self-organizing features like: system degeneracy, nonlinearity or contextual dependency (Dutt-Mazumder et al., 2011; Glazier, 2010; Pedro Passos, Araújo, & Davids, 2012). These studies suggest that self-organization is a functional mechanism that can explain the emergence of interpersonal coordination tendencies within intra-team interactions. However, these prevalent strategies and approaches to the study of team performance are still falling short in providing enough and accurate information about the performance context which is why sports scientists need to rethinking their research strategies (Duarte et al., 2012).

According to Dutt-Mazumder and colleagues (2011) there are mainly two approaches to study sport games as dynamical systems: the application of analytical tools of nonequilibrium thermodynamics, and modeling the dynamics of human movement through the formulation of synergetic and nonlinear equations. As described in this section the latter has received much more attention and has been applied in sports contexts, such as: soccer, badminton, basketball, boxing, rugby union, squash and tennis. In this context it was possible to successfully equate the fluctuations of the systems behavior constrained by the constant perturbations from environment (Dutt-Mazumder et al., 2011). There are however limitations and practical disadvantages in using equations to address (relative phase, dyadic relationships) perturbations as they limited to weak nonlinearities (Beek & Beek, 1988). In order to tackle these limitations Dutt-Mazumder and colleagues (2011) have proposed graphical methods, complex social networks to overcome the difficulty to understand these types of studies and promote a more effective use from practitioners (coaches, athletes). This proposal, has also been followed by several other investigators (Duch et al., 2010; Grund, 2012; Passos et al., 2011; Vilar et al., 2012).

2.2 Complex Network Theory in Team Sports Performance

The simplest working definition of a *network* is probably: "*collection of vertices joined by edges*" (Newman, 2010). Even with this very simple definition one can identify a multitude of domains where network science as shown its usefulness in representing real-world structures like: *communication network* (computers), *social networks* (people), *information network* (World Wide Web), and, *biological networks* (nature) (Easley & Kleinberg, 2010; Newman, 2003). Given this variety, vertices and edges can represent different objects in different domains and even be referred to differently, for example in computer science these are known as *nodes* and *links*, or in physics *sites* and *bonds*, and in sociology *actors* and *ties* (Newman, 2010).

Social network theory is supported by the assumption that seemingly autonomous individuals and organizations are embedded in social relations and interactions (Borgatti, Mehra, Brass, & Labianca, 2009). Therefore, *social networks* are social structures with people (called *actors*), or groups of people, which are related by some form of social interaction (Brandes et al., 2012; Newman, 2010), such as: interdependency (e.g. dyadic interactions between team players and opponents), friendship, kinship, common interest, beliefs relationships, acquaintance, prestige.

Therefore when we consider the study of social relationships under the lens of network theory, we say that we are using *Social Network Analysis* (SNA), particularly if networks are considered explanatory variables and not dependent variables (Brandes et al., 2012). This complexity tool had gained significant relevance in anthropology, biology, communication studies, economics, geography, information science, organizational studies, social psychology, and sociolinguistics.

The network approach to the study of small groups as received a great contribute from (Berkowitz & Wellman, 1988) through the identification of five fundamental principles that provide some underlying intellectual unity to the network approach and that confirms some more recent suggestions and studies under dynamical systems theory framework. These five points are presented as follows:

i. When studying the web of relationships in which people are embedded we will managed to predict better their behavior instead of examining their drives, attitudes, or demographic characteristics. In this way the relationships are themselves constraints to people's behavior.

ii. The focus of analysis should be on the dyadic units, rather than on the units *per si* or their intrinsic characteristics.

iii. Consider how the interdependence among units is assumed, and therefore analytic methods must not assume the independence of the behaviors.

iv. Consider that the interactions (flow of information or resources) between two people depend not only on them but also on each relationship with the rest of the group. This reinforces the assumption that for understanding a social system more is required than just the sum of the dyadic ties.

v. The group existence and formation is dynamical and not perfectly bounded. This means that there are overlapping networks generally having crosscutting relationships to a multitude of groups.

In team sports there has also a growing interest in network studies, considering that both competition and training settings offer an extraordinary opportunity for the study of this phenomena using theses tools as interactions between team members are on display for a large number of events (Duch et al., 2010).

For Duch and colleagues (2010) the team processes that lead to greater performance are still unclear, specifying that there are no clear and definitive answers to how individual strengths and roles (including that of "superstars") are combined for optimal results. According to the same authors, the players' true impact on the team's performance is hidden in the plays of the team. In this study a directed network of ''ball flow'' among the players of a team is used in order to capture the influence of a given player on a match. On the experience of using social network analysis the authors considered it a powerful instrument in order to demonstrate that typical network metrics (such as *flow centrality*, defined later section 3.1.2) provide an objective quantification of individual and team performance.

Similarly to Duch and colleagues (2010), in many studies in sports context that use SNA the interaction types that are considered depend on the ball flow. Typical examples of this interaction occur from the exchange of the ball, between team members, or from their interception from the opponent team players. Therefore, the direction of the interactions (e.g., ball passes) matters and are represented in network science as *directed edges*, and the network in which they are embedded is referred to as a *directed network* or *directed graph* (or even *digraph*) (Easley & Kleinberg, 2010; Newman, 2010). Figure 3 is the graphical representation of a directed network in which the direction of the edge arrows is associated with the ball flow during a volleyball match. The network is obtained by the aggregation of all passes during the match and the edges width and nodes radius is defined by different network metrics.

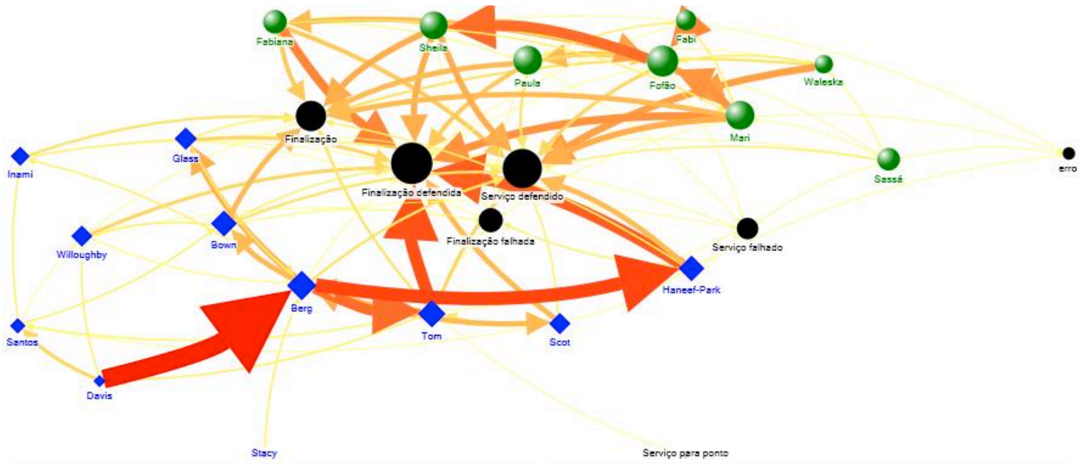


Figure 3. Example of a directed network of the Olympic Volleyball Women Final, 2012 (Source: author, 2012)

In the sports context one can also identify relations between players that do not have a direction, one such example in sports context is considering the goals as pseudo-players in interaction with the players (e.g. mainly the goalkeepers).

If the relation between actors is characterized by symmetry (*undirected networks*) or asymmetry (*directed networks*) in the network is important for a great number of real-world structures like financial, transportation, and as we are seeing, social interactions (Easley & Kleinberg, 2010).

Considering the cardinality of the interactions between a pair of nodes, networks can be simple - *simple networks* or *simple* graph - in the sense they have at the most one edge between any pair of vertices, or with multiedges an therefore called a *multigraph* (Newman, 2010). In the latter, the edges' cardinality can also be represented as having a strength, weight (e.g. number of passes between a pair of players, as in figure 3) or value (normally a real number), which means the amount of interactions/data flowing between the vertices that are linked together (Newman, 2010). However, *weighted networks* do not necessarily result from *multigraphs*, as the weights do not necessarily interaction cardinality, e.g. different distances between cities could be represented by weighted networks, but not necessarily by a multigraph.

Two common metrics that are used to characterize the network are its size (number of edges) and order (number of vertices). An important relation between these two is obtained if we consider the number of edges – network size - (number of passes or actions with the ball) over the number of edges that could exist in the network, we'll have the *density* of the network (Guillaume & Latapy, 2006), representing the intensity of interactions in the social network.

Grund (2012) has focused on the growing interest and evidence in team structural properties of interactions between team members associated to performance outcomes (Balkundi & Kilduff, 2006; Katz et al., 2004), and find out that the analysis focus on team performance must not be on individual team members but on the dyad between team members (the orchestration of interactions and their interactions). So the interactions pattern differences between teams, matters for the explanation of their different rates of success and performance. In his work (Grund, 2012) found that there exists a positive effect on sports team performance from the level of interaction between the team members (network intensity - density), i.e., increases in passing rates are associated to increases in team performance. On the other hand, network centralization lead to a decrease in team performance, i.e., the team has more difficulties in performing well when the team production process is centralized. And this explains why the access to resources and their successful mobilization in tasks that require the involvement of different individuals are facilitated by the relationship between team members.

2.2.1 Connectivity and paths

Considering the interactions between people, it is important to consider not only the immediate and direct connectivity between dyads but also the set of interactions in which they are connected. That is, how nodes can be connected or not via a sequence of links.

In this context, one finds the concept of *walk* which is a sequence of vertices in which each consecutive pair is connected by an edge (Easley & Kleinberg, 2010). Two subsets of walks are also defined: that of *trail* where edges cannot be repeated and *path* where neither edges nor vertices can be repeated (which is not common in team sports). One particular case of a trail are *cycles* that are represented by a "ring" structure, repeating only the first and the last nodes (Easley & Kleinberg, 2010); (Newman, 2010).

The most relevant metric over paths is their length. This is on unweighted graphs obtained by counting the number of edges in the path. On the other hand, on weighted graphs this is typically obtained by summing the weights of the edges along the path. A related concept is that of *shortest path* or *geodesic path*, that represents the path between two vertices for which there is no other path in the network that is shorter than that path. The length of the shortest path, is the *geodesic distance* or *shortest distance* that represents the shortest network distance between those two vertices (Newman, 2010).

Paths are not common in sports (see Ramos, Lopes, & Araújo, 2017), therefore models such as *random walks* (a walk that takes random steps across the network) may represent better the events during a team sport match, such as ball passing (Newman, 2010). Both concepts of length and random walks can be enlightening if associated to performance outcomes, once there is some debate (from practitioners) about its relation to the means to score goals, whether if is preferred the use of 'longer passing' in soccer or 'direct play' rather than for 'possession play' (Vilar et al., 2012). This presumed correspondence remains unclear because of a lack of a theoretical understanding (Vilar et al., 2012).

2.2.2 Nodes' Centrality

A large volume of research on network theory has been devoted to the *centrality* concept (Newman, 2010), which in sports context, means to try to answer questions like: "which are the most important or central actors (players) in a network (team)?".

The simplest measure of centrality is the *degree* of a vertex (also called *degree centrality*), and obviously is applied to the vertex (player) and is defined by the number of edges (interactions) connected to it (Newman, 2010). This metric for vertex *degree* is applied typically to undirected networks; although, as previously mentioned, team ball sports have been studied mostly looking at the "ball flux" (e.g., ball passes) which are directed networks and therefore vertices (players) have two degrees: the *in-degree* (the number of ingoing passes to that player, or interceptions made) and the *out-degree* (the number of the outgoing passes or interactions conceded) (Newman, 2010).

The centrality analysis can also be conducted by another metric in cases when a connection to a popular individual is more important than a connection to a loner. The *Eigenvector Centrality* metric takes into consideration not only how many connections a vertex has (i.e., its degree), but also the degree of the vertices that he is connecting to

(Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006; Easley & Kleinberg, 2010; Newman, 2010). Typically one of the questions asked, which can be cleared out by the use of this measure, is: “Which node in this network would prove most crucial to the network’s interactions if she/he was removed?” (Newman, 2003).

Another commonly used concept and metric for centrality is *Betweenness centrality* (Boccaletti et al., 2006; Easley & Kleinberg, 2010; Girvan, 2002; Newman, 2003, 2010) and it is obtained by counting the geodesic paths between vertices that run along each vertex in the network. As previously mentioned the node degree can give some insight about the importance or popularity of someone in the network. Although this metric does not take into account the “bridging” function that can be performed by an actor that is connecting other popular or important roles in a social network. If this (connecting) person were removed from the network, those other important nodes would be disconnected from each other. A node that has an important bridging role can be identified by having high Betweenness Centrality. On the other hand, if some node were removed from the network, and everyone would still be connected to everyone else and their shortest path would not even be altered than that player as Betweenness Centrality of 0. In short, vertices that are included in many of the shortest paths between other vertices have a higher Betweenness Centrality than those that are not included. This centrality measure can also be viewed, as a measure of network *resilience*, indicating how much effect on path length the removal of a vertex will have. For Newman (2003), networks vary in their level of resilience according to such vertex removal.

On the other hand, how close each person is from others in the network can be accessed by another metric for centrality called *closeness* (Boccaletti et al., 2006; Newman, 2010), which is the inverse of the average distance from all other persons. Considering that information flowed through edges in the network, some people would be able to contact all the other people in only a few steps, while others may require many steps. Unlike other centrality metrics, a lower Closeness Centrality score indicates a more central (i.e., important) position in the network, which corresponds to the lowest Closeness Centrality measure, suggesting that she/he may be in a good position to spread information/influence through the network efficiently.

2.2.3 Group formation: cliques and clusters

To a better understanding on what clustering is, we first must understand other relevant properties of the networks like *transitivity*, which “in the language of social networks, means that the friend of your friend is likely also to be your friend” (Newman, 2003, p.183). This closure of a triangle of relationships, accordingly Easley and Kleinberg (2010) is called *triadic closure*, and the presence of a heightened number of triangles in the network is the transitivity of the network, which can be quantified by defining the *clustering coefficient*. We can say in a simple way that the Clustering Coefficient measures how connected a vertex’s neighbors are to one another. Therefore, clustering coefficient measures the triangles *density* in a network. The term *reciprocity* is often measured in directed social networks, to obtain the probability that two vertices point each other (Newman, 2003).

On local clustering, the authors (Guillaume & Latapy, 2006) refers that all real-world complex networks exhibit a high clustering and it appears to be independent of the size of the network.

Therefore we’re talking about a network area that concerns groups of vertices, like *cliques* or *plexes* (used for discovering groups within groups) and *cores*. This is an evolving area since in the last decade and there’s been a enormous increase in research on social networks and small groups (Katz et al., 2004). Therefore and from the network theory perspective in the social network literature, a group could be defined in two different ways. First, as a structural feature of a network, which means that within some population there are emergent subsets (cliques) of fully connected (or almost) nodes (Katz et al., 2004). One problem related to the identification of this cliques is the required criteria of cutoff values, but Freeman (1992) applied the idea of strong and weak ties to distinguish those subgroups. Second, a group is externally categorized or bounded like players in a team, students in a class, or a corporation.

2.2.4 Network models: small-world and power-laws

The research on graph theory focused not only in graphs but also on classes of graphs, creating what is known as theoretical network models. Most of the theoretical models are random and created through stochastic processes (Rocha, 2011). The formal proposal that popularized random models is due to Erdös and Rényi (Erdös & Rényi,

1959) and the ER model differs from others based on fact that the vertices are connected to a fixed number of other vertices or with a fixed probability. The ER model was developed after Solominoff and Rapoport (1951) works on random networks (Solomonoff & Rapoport, 1951). However the properties of these random structures are not representative of empirical ones (Rocha, 2011) and other models have been created after that.

One such phenomenon is called the *small-world effect* (Boccaletti et al., 2006; Easley & Kleinberg, 2010; Watts, 2003) that reflects the idea that the world is small and everyone is somehow connected to each other. This idea came from the six degrees of separation, famously expressed by a phrase of a play from John Guare, referring to Jeffrey Travers and Stanley Milgrams letter experiment (Travers & Milgram, 1977), with that title (Easley & Kleinberg, 2010, pp.35): "*I read somewhere that everybody on this planet is separated by only six other people. Six degrees of separation between us and everyone else on this planet.*" Watts and Strogatz developed the *small-world model* (WS model) to represent the many closed triads, but also very short paths, which is found in many real-world random networks that reveals the referred small-world effect (Easley & Kleinberg, 2010; Watts, 2003). The scope for this model is associated to large networks, in which nodes can be linked through only a few links (i.e., small geodesic paths) and their applicability to small groups was not yet empirical demonstrated.

On the other hand, in team sports context, Passos and colleagues (2011) founded something similar to the *small-world effect* in the interactions of small units of system agents that originated their interest. These authors consider that few vertices (players in games sub-phases) maintain connection through a path of few links, and this is like interactions of small units of team players with opposite team players in sub-phases competitions. In their study in team sports (Passos et al., 2011) could confirm the usefulness of the small-world network concept in capturing pattern formation dynamics.

Passos and colleagues (2011) also pointed out that the *preferential attachment* property could reveal some important information about the mode of control that is adopted in different team game performance contexts. Meaning that, if one team exhibits a fixed pattern of specific preferential attachment between few players, it becomes predictable and therefore the other team can perturb the preferential interactions between those players. The preferential attachment is ''the rich get richer'' paradigm, because the most connected vertices have higher probability of receiving new vertices (Easley & Kleinberg, 2010). Barabási and Albert (1999) showed that many real systems are characterized by an uneven distribution and their network model (AB model, *scale-free networks*) uses the "preferential attachment" principle as an explanation for the power law degree distribution. This is, characterized by vertices being highly connected while others have few connections. In this kind of networks there are vertices that are linked to a large amount of the edges of the network, and thus called *hubs*.

About the distance between two vertices (the number of edges on a shortest path between these vertices) we've already referred the property of social networks the "six degrees of separation", that helped to understand the low average found in this real-world complex networks (Guillaume & Latapy, 2006).

Returning to Guillaume and Latapy (2006) proposed model and to the unexpected results on the degree distribution that follows a power law, this means that there's a significant number of vertices with high degree and the other majority have a small degree. In their study both the WS model and the AB model were introduced to model generic behavior of complex networks, failing (both) in producing graphs having each of the three properties we cited. The WS model gives a possible explanation for the high clustering of complex networks that is the locality of the links and the AB explained the power law distribution with the preferential attachment principle. Therefore justifying the new model proposed.

This work had been clarified in some of the directions that can be taken, given that the model haven't been tested on directed and weighted graphs (those that exists in team sports performance studies with SNA), and also revealing some weakness on the definition of cliques.

2.2.5 Bipartite networks and hypernetworks

When an edge join more than two vertices at a time, like in family ties, that edge is called a *hyperegde* and the corresponding network a *hypergraph* (Fig.2) or a *hypernetwork* (Johnson, 2013; Ramos, Lopes, et al., 2017a).

In sociology this networks are called *affiliation networks* and they can also be represented as a *bipartite network* (Fig.3). This *two-mode network* (also called this way in sociology literature) represents two kinds of vertices that are not connected among equals, just between the other types of vertices. A example is to consider a network of players that participates in one team game, and the bipartite network that represent it, has this two types of vertices: players and the technical skills that each have done, connecting only by an edge each player to the each skill that have done (Newman, 2010).

In a different context than team sports performance, but with some interest to it, in a study on bipartite graphs (Guillaume & Latapy, 2006) showed how all complex networks may be described as bipartite structures, presenting a model that can be tested for any kind of real-world complex network. In their studies they pointed out a model where some properties are common to all complex networks. These properties are: the low density of the network, the average distance between vertices, the high clustering and the power law degree distribution.

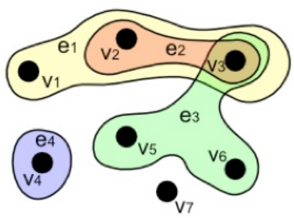


Fig.4 Sample of hypergraph, with X = {v1, v2, v3, v4, v5, v6, v7} and E = {e1, e2, e3, e3} = {{v_1,v_2,v_3}, {v_2,v_3}, {v_3,v_5,v_6}, {v_4}}. Taken from: http://en.wikipedia.org/wiki/Hypergraph

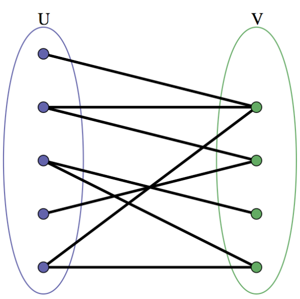


Fig.5 Example of a simple bipartite graph. Taken from: http://upload.wikimedia.org/wikipedia/commons/e/e8/Simple-bipartite-graph.svg

One interesting feature that results from the projection of the bipartite network in the two one-mode projections, one for each group of vertices, is a cluster of vertices that are all connected to each other, which means a *clique* (in network jargon) (Newman, 2010).

2.2.6 From origins to the state-of-the-art in SNA

We have seen how SNA as been applied in some studies in PA, and it may be interesting now go back a little to its origins and understand some of its evolution.

In short, the origins of the field of social networks are related to sociologists who have the longest tradition in quantitative and empirical work in this field. It can be found remote literature from the end of nineteenth century, but the real foundation of the field is attributed to psychiatrist Jacob Moreno (Newman, 2010), a Romanian immigrant to America, who started is interest in the dynamics of social interactions within groups of people, in the late 1930s. This researcher seeded the origins of *sociometry*, through what he called human interaction *sociograms* published in his book "*Who Shall Survive?"* (Newman, 2010). Social scientists were easily persuaded because once one draws a picture of a network it's easy to see and are also sociological interesting.

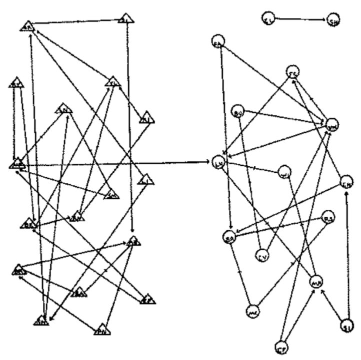


Figure 6: Friendships between schoolchildren. Hand-drawn image of a social network, taken from the work of Moreno, depicts friendship patterns between boys (triangles) and girls (circles) in a class of schoolchildren in the 1930s (Newman, 2010, pp 37).

The most common method for determining the structure of social networks is the *direct questioning* of the people. However *archival records* are another important technique, like the one made in the 1030s ("Southern Women Study", Davis, 1941), in the US, through newspapers data about the appearance of 18 women in 14 social events (Newman, 2010). These women would be considered connected if they appeared in the same event, creating an *affiliation network* or a *bipartite graph*.

The powerful properties of networks have since Moreno and Davis et al. works created such a development and interest that had been applied to a large variety of different issues and problems. Direct questioning or questionnaires were helped by the use of computers and the use of online survey tools, until than studies were limited to a few tens or sometimes hundreds of respondents.

These social sciences applications began in the 1950s, when in sociology and anthropology the quantitative methods were being more applied and this mathematical language of graph theory were also helpful in understanding the data from those studies at that time (Newman et al., 2006). Similarly at this time, the propagation of information and diseases were being seeing as graphs by mathematicians (Newman et al., 2006). Some heritage resulted from this both areas of interest, for social network analysis, and from the social sciences came most of the terminology used, like: path lengths, actor centrality, cliques, connected components in matters like status, influence, cohesiveness, social roles and identities in social networks (Newman et al., 2006), and also became a practical tool for the analysis of empirical data. On the other hand, and from the mathematicians applications, it was more the behavioral characteristics like the estimation of the size of an epidemic or even the possibility of global information transmission, that were brought by the structural properties of networks (mainly connectedness). The stochastic property of objects associated to graphs, though in terms of probability distributions, is also another heritage of that time, and accordingly Newman and colleagues (2006) with great deal of interest.

Therefore, the evolution of networks stepped into:

i) The concern with empirical work and theoretical questions, focusing on the properties of real-world networks;

ii) How dynamical feature evolve through time;

iii) Considering networks as a framework where distributed dynamical systems are built upon, and not only topological objects.

In summary, traditional theories of networks did not pay much attention to the structure of naturally occurring networks, especially to networks arising in the real world, concerning more with artificial constructs; on the other hand SNA, tends to be more descriptive rather than constructive, because of its strongly empirical feature, avoiding modeling and preferring simply the description of the properties of the collected data (Newman et al., 2006).

Additionally, the static features of the networks, considered both in graph theory and SNA are not adequate (Albert-Laszlo & Reka, 1999) and it has been clearly shown that networks evolve over time (Watts, 2003). Furthermore, most of the social networks are in fact, the product of dynamical processes that constantly add or remove edges or vertices, and consequently evolving dependently from the role of the participants and their emergent pattern of behavior (Newman et al., 2006).

Also, the traditional approaches to networks overlooked and oversimplified the relationship between the structural properties and the behavior of the networks (Newman et al., 2006). On the other way, some significant amount of more recent work (Duarte et al., 2012; Duch et al., 2010; Grund, 2012), approached with a dynamical systems view, representing dynamical entities by the vertices of the graph with their own rules of behavior, and the couplings between the entities represented by the edges, finding not only topological properties, but also dynamical properties (Newman et al., 2006).

In this thesis work, we have started from this point (SNA), and moved to a complex networks approach, as described in previous section 2.2.5. The following chapters describe, via a set of core papers, our contribution in applying complex systems approach to PA in team sports context.

3 Core Papers

3.1 What's next in Complex Networks? Capturing the concept of attacking play in invasive team sports

3.1.1 Context and summary

In this review article, we propose a novel approach to team sports performance centered on sport concepts, namely that of an attacking play. Network theory and tools including temporal and bipartite or multilayered networks were used to capture this concept. We aimed to leverage the understanding of the structure and dynamics of invasive sports teams related to their performance.

We start, in section 2.1, by discussing how SNA has been commonly used to address team structure and dynamics in invasive team sports performance. Typically this means aggregating all the passes between players in a single directed network for which measures are taken and related to success indicators (e.g. reaching a competition stage, goals scored) (Duch et al., 2010; Fewell et al., 2012; Grund, 2012; Travassos et al., 2016). In section 2.2, we describe temporal bipartite networks in depth and the way they may overcome the shortcomings often found when applying SNA to PA. We analyze the attacking play concept introducing nodes that represent other players' actions, instead of passes. The chief contribution is that these nodes are part of a bipartite network, which allows us to retain the key concept of attacking play during network analysis and relate it to its possible outcomes. In section 2.3, a set of questions that are commonly found in the literature are placed as illustrative examples of how our suggestions can be used in match analysis in a way distinct from previous studies. Although we do not address directly either players' position or the interaction between both teams, we present, in sections 2.4 and 3, guidelines for future studies in this direction.

We put forward eight questions directly related to team performance to discuss how common pitfalls in the use of network tools for capturing sports concepts can be avoided.

We propose that temporal and bipartite networks could be an alternative approach for representing the interactions between players during a game. Using the flexible time structure of temporal networks it is possible to capture the sequence of passes in an attacking play, which is one of the main concepts of team collective behavior. We have highlighted how temporal bipartite network representation empowers existing metrics for capturing sports fundamental concepts (e.g. style of play) with greater adequacy. Moreover, we suggest that methods combining spatial and hypernetworks (Johnson & Iravani, 2007) with temporal networks represent a promising direction for future research, as they allow the analysis of dynamics of the network. These complex networks could integrate concepts such as how time changes the structure of the network, as well as the players' technical resources and their positioning relative to the position of other players (team-mates or adversaries).

Finally, we propose that, at this stage of knowledge, it may be advantageous to build up from fundamental sport concepts toward complex network theory and tools, and not the other way around.

3.1.2 Paper *author* copy

What's next in Complex Networks? Capturing the concept of attacking play in invasive team sports.

João Ramos [1, 2], Rui J. Lopes [1, 3] & Duarte Araújo [4]

[1] ISCTE-Instituto Universitário de Lisboa, Lisbon, Portugal

[2] Universidade Europeia, Laureate International Universities, Estrada da Correia, 53, 1500-210 Lisboa, Portugal

[3] Instituto de Telecomunicações, Lisbon, Portugal

[4] CIPER, Faculdade de Motricidade Humana, Universidade de Lisboa, Cruz Quebrada Dafundo, Portugal

Key Points

Network theory can contribute to performance analysis in invasive team sports by describing the complex and dynamic features of interaction between players.

Complex networks, notably temporal and bipartite networks, can capture the concept of attacking play by including play actions and their temporal sequence.

Typical questions on the team interaction properties can be answered by applying, possibly at different scales, wellknown bipartite network metrics, yielding different results from usual social network analysis.

Abstract

The evolution of performance analysis (PA) within sports sciences is tied to technology development and practitioner demands. However, how individual and collective patterns self-organize and interact in invasive team sports remains elusive. Social network analysis (SNA) has been recently proposed to resolve some aspects of this problem, and has proven successful in capturing collective features resulting from the interactions between team members as well as a powerful communication tool. Despite these advances, some fundamental team sports

concepts such as an attacking play have not been properly captured by the more common applications of SNA to team sports performance. In this review article, we propose a novel approach to team sports performance centered on sport concepts, namely that of an attacking play. Network theory and tools including temporal and bipartite or multilayered networks were used to capture this concept. We put forward eight questions directly related to team performance to discuss how common pitfalls in the use of network tools for capturing sports concepts can be avoided. Some answers are advanced in an attempt to be more precise in the description of team dynamics and to uncover other metrics directly applied to sport concepts, such as structure and dynamics of attacking plays. Finally, we propose that, at this stage of knowledge, it may be advantageous to build up from fundamental sport concepts toward complex network theory and tools, and not the other way around.

1. Introduction

The evolution of performance analysis (PA) as a sub-discipline of sports sciences has seen significant advances in team sports. Technological progress has had an important role in this process, which is reflected in the availability of more data and with better accuracy and higher precision (1). The data processing phase of PA has also evolved from a mostly descriptive and qualitative approach to a quantitative and complex software-based analysis. PA approaches began with *biomechanics* and *notational analysis*, which were centered on individuals and their positions, actions and time (2-4). Notational analysis is based on the quantification of critical events through frequency counting followed by qualitative and quantitative feedback analysis. First, the performance indicators for the evaluation are determined (mainly on-the-ball actions of players), then the detail level is selected manually or computationally, and finally extensive flowcharts are created for the analyzed performance indicators (4). This quantification of critical events follows a sequential formula (player → position → action → time) normally associated with success (3-5). The biomechanics approach mainly entails individual fine-grain analysis and even when it is applied to team sports, it focuses on individual technique performance (3). This method has been mostly used in the context of closed sports techniques, particularly in performance environments (3). Aspects such as the non-linear interactions between the many independent parts of emergent movement patterns, or the interactions between players or even between teams, clearly show that theoretical guidance is essential for understanding complex adaptive sociobiological systems like team sports (2). A single event such as an injury at the microscopic level in a soccer player can have a large-scale impact on his/her performance, and an even greater effect on the team's structure and team's performance (micro-macro relatedness). A comprehensive PA therefore requires interdisciplinary approaches complementing biomechanics and notational analysis. Several authors (2, 6-8) believe that dynamical systems theory (DST) gathers the necessary theoretical framework conditions for PA because it takes into account the stability, variability and transitions among interaction states, including at different levels of analysis, and compares them with specific outcomes. Indeed, self-organization explains how order emerges from the interaction of different components; for instance, how an individual performance results from the interaction of the player's body segments, or how team performance results from the interaction of individual team players (9).

Following such innovative approaches, the next logical step in PA is to address complex issues such as the relationship between team structure and dynamics, as well as the types of team interactions and their interdependency within and between teams (1-3, 5, 10, 11). In particular, it is important to understand how players' actions can disrupt the equilibrium of the other team or create scoring opportunities. Dutt-Mazumder and colleagues (6) proposed more visual methods to overcome the complexity of these issues and to encourage practitioners (coaches, athletes) to use these approaches. In a similar way, sports scientists have utilized social network analysis (SNA) (5, 12-15); notably, the communicative power that network visualization offers for new insights into network structures (16).

Concerning the dynamic features and their study, in network theory, two different types of dynamics are usually considered: dynamics *on* the network and dynamics *of* the network (17). Dynamics on the network focuses on flows across the network structure, e.g. ball passes between players. On the other hand, dynamics of the network are concerned with changes in the network structure itself, e.g. how the players' positions can provide information on a possible pass or player action. In the current review, we focus on the first type of network dynamics. Successful applications of dynamic network visualization in the literature mostly represent the "ball flux" in static *flip books*, where node position remains constant but edges cumulate over time, and therefore the researcher obtains cumulative snapshots of the network as a function of time (18). Typically SNA applied to PA uses only the ‘last snapshot’, that is, the network resulting from the aggregate of all the interactions occurring during the entire match; focusing on the structure and not so much on the dynamics. We believe that this impairs not only the study of relevant dynamic features of the team, but also conceals import concepts such as that of an attacking play. Dynamic relations in team sports games have other basic dimensions, in addition to passes, that have not yet been fully captured in sports settings: i) relational space (i.e. interactions considered in a geographical space); ii) their time structure (i.e., rate change, order or sequence, or simultaneity of interactions); and iii) their relations with types of nodes (i.e. colleagues or adversaries), meaning cooperation or competition interactions (18), as discussed in the next section. This is a possible approach for handling complex adaptive systems including team sports analyses in competition. Thus, network dynamics analyses may explain how and why disequilibrium situations such as scoring opportunities occur.

In this review we aimed to leverage the understanding of the structure and dynamics of invasive sports teams related to their performance. In particular, we focused on the interactions between players in the same team (ball passes), their actions during the attacking phase (e.g. ball recovery, shot at goal), and the temporal structure of both. We did not address directly either the players’ position or the interactions between both teams. We start, in section 2.1, by discussing how SNA has been commonly used to address team structure and dynamics in invasive team sports performance. Typically this means aggregating all the passes between players in a single directed network for which measures are taken and related to success indicators (e.g. reaching a competition stage, goals scored) (12, 14, 15, 19). In section 2.2, we describe temporal bipartite networks in depth and the way they may overcome the shortcomings often found when applying SNA to PA. We analyse the attacking play concept introducing nodes that represent other players’ actions, instead of passes. The chief contribution is that these nodes are part of a bipartite network, which allows us to retain the key concept of attacking play during network analysis and relate it to its possible outcomes. In addition, by applying temporal networks, dynamic aspects of the team process can be captured, namely the time structure of the different attacking plays. Both concepts, bipartite and temporal networks, can be extended hierarchically, enabling performance analysis within different event and time scales. In section 2.3, a set of questions that are commonly found in the literature are placed as illustrative examples of how our suggestions can be used in match analysis in a way distinct from previous studies.

Although we do not address directly either players' position or the interaction between both teams, we present, in sections 2.4 and 3, guidelines for future studies in this direction. Notably, in order to use spatio-temporal structure and dynamics of both teams (represented by players' positions) in the analysis of the team and players' actions, the temporal structure of such actions must be adequately represented. Temporal networks can provide such representation. On the other hand, the outcomes of these dynamic interactions are naturally represented by different layers in bipartite networks.

2. Network Theory and Tools in Performance Analysis

2.1 Why use Network Analysis for Performance Analysis?

Has PA succeeded in understanding the processes, such as structure and dynamics, leading to improved performance? In this article we have raised some questions and problems that currently challenge the research in the field. Specifically, we ask what is the importance of a player in the structure and dynamics of the team/network, besides providing individual performance indicators? And who is the most connected player (i.e. playmaker) regarding his/her number of colleagues and interactions? Finally, what team sub-units, such as pairs or triangles of players, have the strongest influence on team performance? These and other questions, in particular those addressing team structure and dynamics, can be investigated using networks theory and tools such as SNA. However, we do not aim and it would not be possible to comprehensively review the applications of SNA to team sports performance. Instead, we will discuss the limitations of SNA for explaining team performance and how they may be overcome. In our view, until these issues are fully resolved SNA will be insufficient for clearly describing and explaining the dynamic nature of the processes associated with team performance. We propose the adoption of multilayer networks, in particular, of bipartite and temporal networks, to overcome some of the limitations of SNA. To achieve this aim, we present solutions from network theory for accurately representing dynamics and suggest alternative metrics that we find more adequate for explaining sport teams performance.

The definition of *network* is a "collection of vertices joined by edges", which can represent the pattern of connections between different objects (20). Thus, social networks are structures with persons or sometimes groups of persons (actors) who are related by some form of social interaction (ties) (20, 21). In team sports settings, we can consider the actors as the players, and the ties as the interactions between players. Team goals are achieved if each individual's effort is coordinated with those of the other team-mates through dynamic interactions, i.e., a complex network is considered rather than the simple summation of the individual performances (2, 5, 19, 22, 23). To address these team sports characteristics, we focus on eight questions that relate directly to team performance and network metrics.

Technical terminology and heavy jargon from complex systems approaches can be overcome by using network visual representations, which are powerful and versatile tools widely used to describe dynamical systems (5, 6, 12, 16, 18, 24, 25). Typically, in SNA applied to sport matches, the nodes represent team players and the links between nodes correspond to the interaction between those players (19), specifically "ball passing"[2]. This relationship is characterized by a *transport action* (i.e. a token - ball - is passed between players) and a *directionality* (i.e. for each interaction there is a sender and a receiver), which combined make a directed network (i.e. the links have a direction).

The two most commonly used systems for representing a network are matrices[3] (adjacency or incidence) and graphics. While matrices are particularly relevant and useful for the development of formal methods and computational processes, graphics have an extraordinary communicative power. Notably, through the effective use of network graph visualization features, such as the nodes' and edges' position, size and colour, particular properties of the network, i.e., the importance of a player or frequency of an interaction, can be highlighted and perceived intuitively without requiring a specialized knowledge of network theory. For example, to identify the number of passes

---

[2] Analyzing only at ball passing restricts the analysis of team performance to the attacking phase. In the current article, we do not attempt to directly resolve this limitation.

[3] An adjacency matrix, A, is a square matrix, with rows and columns representing nodes (e.g., players) with entry a_{ij} of *A* taking value 1 if there is a link between node *i* and node *j*; and 0 otherwise. Different types of networks lead to different matrix structures: undirected graphs are represented in symmetric adjacency matrices, the fact that the link between nodes i and j has no directionality is expressed in equality $a_{ij} = a_{ji}$; in directed graphs (or digraphs) the links between nodes have a directionality; a link from node i to node j is expressed by entry a_{ij} taking value 1 independently of the value of a_{ji}. In this paper the links represent actions by the players (e.g., making a pass) and are thus directed leading to digraphs.. In what are called weighted graphs, the entries of the matrix can take other values w_{ij}, called weights, that are nor restricted to 0 or 1. The value taken by entry w_{ij} reflects the intensity or strength of that link.

In an incidence matrix, E, rows represent nodes and columns represent links. The entry e_{ij} takes value 1 if the link *j* is incident on nodes *i* and *j*; 0 otherwise. In directed networks values -1 and 1 are used to distinguish link origin and destination.

by each player, instead of summing matrix values, the performance analyst can simply observe the size of the nodes (to identify the most connected player) or the weight of the edges that represent the number of passes between two players (identifying the dyadic that interacted more). This intuitive characteristic of network graphs and the non-specialized skills required from the reader are being explored by generalist newspapers in articles devoted to soccer match analysis. Figure 1 uses one of these generalist newspaper examples to illustrate some of the features mentioned above (26).

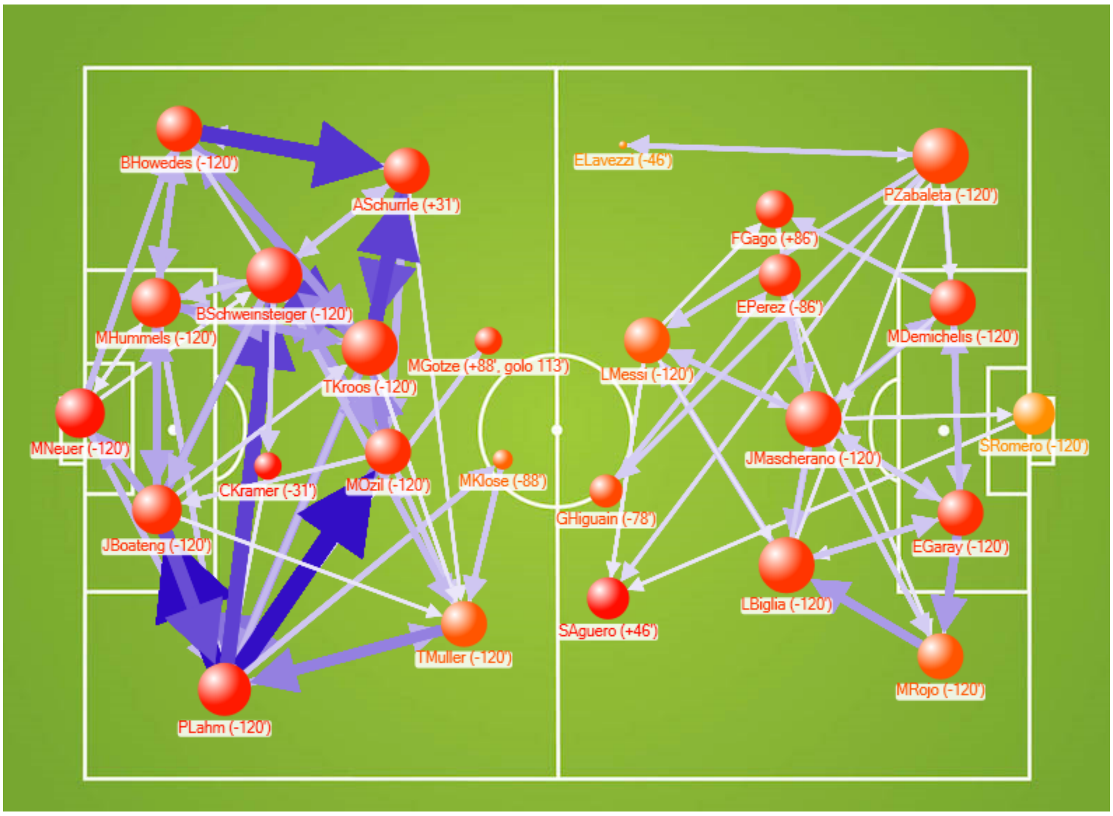


Figure 1: Pass interactions in the Germany vs. Argentina match for the 2014 FIFA World Cup (26). Each circle represents a player in his relative position; the radius and colour of each circle represents the number of players that player interacts with and his pass precision (red more precision and yellow less precision), respectively; and the arrows represent the direction of the passes between players; the width and shade of each arrow represents the number of interactions (passes) between players (lighter arrows indicate less passes, darker arrows indicate more passes). The numbers in brackets represent the minutes played by the player (e.g., -120', played 120 minutes) or the moment in the match when the player started to play (e.g., +88', golo 113', entered in the match at minute 88 and scored a goal at minute 113).

2.2 From static, single-layered networks to temporal bipartite networks.

Graphs as illustrated in Figure 1 present features associated with nodes and links that are cumulative (i.e. they aggregate all the interactions that have occurred during the match) but not the relevant events in the match such as goals and gaining/losing ball possession, which are not usually represented in these analyses. However, although global metrics can be obtained in this type of network representation (e.g. which player does more passes), the fundamental concept of attacking play is not apparent. An attacking play is defined as the tactical situation when one team is in possession of the ball moving towards the opponent's goal (e.g., Lucchesi [(27)]). To visualize an attacking play it is essential to include in the network other types of nodes to illustrate the beginning and end of the attacking play, such as gaining/losing ball possession, gaining a free-kick, and scoring a goal. However, adding these nodes to graphs in a simplistic manner not only breaks the semantic homogeneity of nodes and links, which do not always correspond to passes between players, but also changes the metric values of the network. We propose that multilayer networks can bring an important contribution to the understanding of team attacking dynamics, specifically, through the combination of temporal and bipartite networks.

Typically, the formalization of a temporal network starts with the definition of M time snapshots for the entire duration of the game, T, with equal interval $\tau = T/M$, and the N nodes of interest. The next step is the aggregation of all the interactions that occur in the time interval (or time snapshot) t, between $(t-1)\tau$ and $t\tau$ where $t = 1, \ldots, M$ (28). Fixing a value for τ not only raises the problem of allocating it an appropriate value but also does not guarantee that the concept of play is accurately represented. We suggest that each interval in the temporal network should correspond to the duration of an attacking play. The definition of the beginning and end of each of these intervals is therefore defined by the beginning and end of an attacking play, typically corresponding to those instants when ball possession has been won or lost. Formally, time snapshot t_i, corresponding to the i^{th} attacking play, is defined by the time interval bound by instant t_{Gi}, where the team gained ball possession and instant t_{Li} where possession was lost.

The introduction of these new nodes to temporal networks takes us to the second element of our proposal: bipartite networks. In bipartite networks, also known as two-mode networks, two different types or classes of vertices are considered, and nodes of the same type cannot be connected directly. Consequently, the links are always incident on nodes of different types. A possible application of the bipartite network model could consider as the two node classes i) the set of players; and ii) the set of technical skills or play actions performed by the players. The application of the temporal and bipartite network concepts in combination is illustrated in Figure 2.

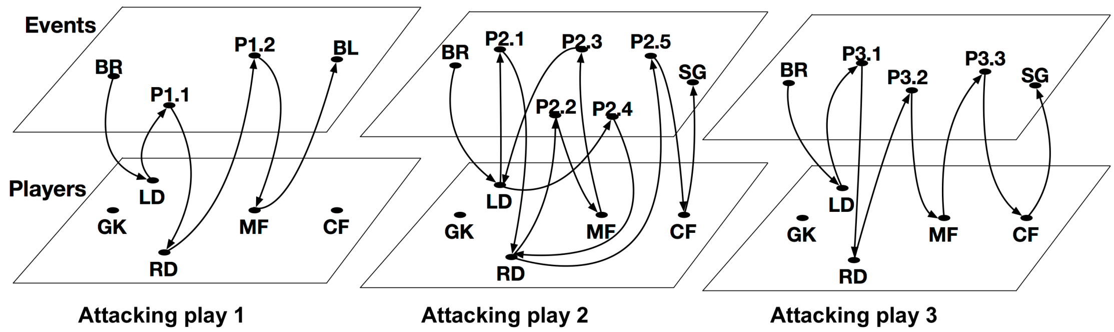


Figure 2: Multilayer representation of three attacking plays in a bipartite (two-mode) network. GK goalkeeper, LD/RD left/right defender, MF midfielder, CF center forward, BR/BL ball recovery/loss, P pass, SG shot at goal. The first number represents the number of the play and the second, the order of the event.

Therefore, the use of any kind of team performance representative nodes (e.g. shot at goal) allows a direct way to relate PA metrics and network-based metrics associated with the attacking play structure. For example, the in-degree metric for the shot at goal node provides directly the number of shots at goal. On the other hand, the average length of the walks leading to this node provides the average number of passes until a shot at goal is made and thus can provide an hint about the team's attacking style of play.

An interesting feature that results from projecting the bipartite network into the two one-mode projections is illustrated in Figure 3.

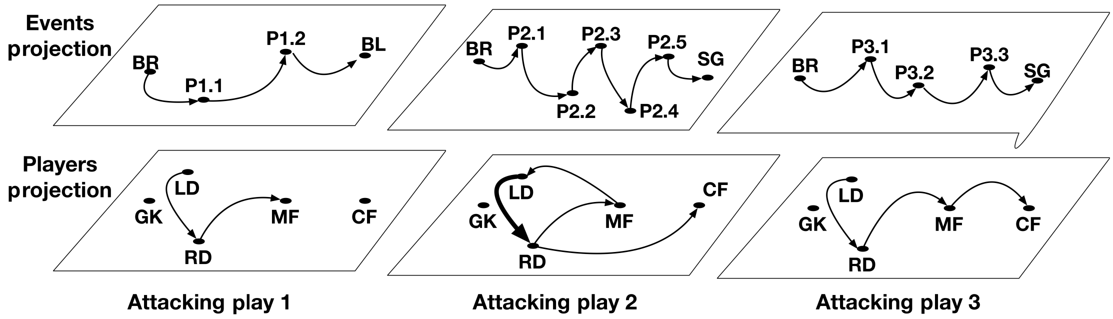


Figure 3: Projection of the multilayer graph in two single-layer graphs. GK goalkeeper, LD/RD left/right defender, MF midfielder, CF center forward, BR/BL ball recovery/loss, P pass, SG shot at goal. The first number represents the number of the play and the second, the order of the event. The thickened arrow in attacking play 2 represents the aggregation of events/passes from LD to RD.

These two projections focus on different aspects of the team performance and uncover important limitations of traditional network match presentations. The events projection reveals the pass path between gaining and losing ball possession. The players' projection in Figure 4 shows the interactions between players (passes) in a similar way to the network in Figure 1.

It is worth highlighting the network concepts that emerge from both projections: i) in the events projection the result is always a path (no node is visited more than once); ii) in the players projection the result is a walk, where a node can be visited more than once (right defender [RD] in the example represented in figure 4). Moreover, team characteristics can be observed including a cluster of vertices that are all connected to each other and are known as a 'clique' in network jargon (e.g. cluster formed between left defender [LD], RD and Midfielder [MF] in figure 4) (20). In the events and players projections one can perform an aggregation operation as shown in Figure 4 (aggregation of a play showing a shot at goal).

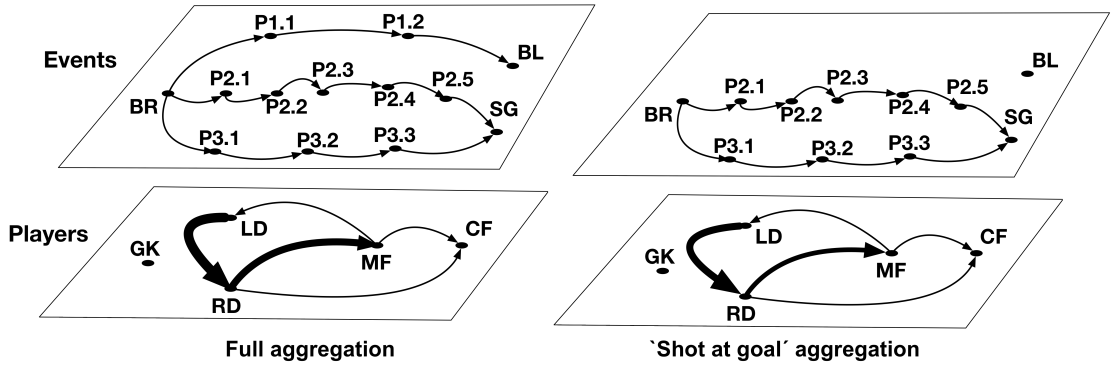


Figure 4: Temporal aggregation of events and player projections. GK goalkeeper, LD/RD left/right defender, MF midfielder, CF center forward, BR/BL ball recovery/loss, P pass, SG shot at goal. The first number represents the number of the play, and the second represents the order of the event. The thickened arrow represents the aggregation of events/passes from LD to RD and from RD to MF.

In addition, some filtering can also be applied to the aggregation process, for instance, to consider only the attacking plays that ended in a 'shot at goal' event, as represented in Figure 4. The projections in Figure 4 clearly show that the concepts of 'path' and 'walk'(29) are of great relevance, as they uncover another distinctive aspect between team sports and other networks, specifically, how each node establishes the outbound link when a token (i.e. a ball in team sports such as soccer) is received. Typically in an invasive team sports match, the trajectory along the network nodes does not follow the shortest path nor even a path where neither nodes or links are repeated (obtained over the full aggregation graph), as is often assumed in sports sciences (15, 30-32). However, as illustrated by Figure 3 (attacking play 1), in many networks the trajectory can be a trail (where links cannot be repeated) or a walk (nodes and links can be repeated with no restrictions), and is not deterministic. This fundamental difference has a very strong impact in many networks metrics, centrality in particular, as discussed in sections 2.3.2. and 2.3.3..

2.3 Using temporal bipartite networks: some illustrative questions.

2.3.1. Question 1: who is the most interactive player?

The activity level of a player in his/her interaction with other team-mates can be captured by the *node degree*[4] of a vertex in a given graph, in which case only the

[4] *Degree of a vertex* i, hence v_i, is given by the number of nodes that are directly connected with the focal node;

adjacencies of the node/player are considered; this is therefore a local analysis of node centrality. In sports settings, the node centrality is the most widely used network metric (12-15, 19, 33) and it is typically based on ball flow. Given the directed nature of the interaction, the degree centrality is divided into two categories: *in-degree*, which measures the number of players who pass the ball to the focal player, and *out-degree*, which measures the number of players to whom the focal player passes the ball to. As centrality is focused on each individual player and his/her participation in the team ball passing activity, it can be represented for the duration of the entire game in simple graphs (Figure 1). The approach proposed in this article (section 2.4) extends the reach of this metric in two ways: i) by defining different time spans for the aggregation process, the activity of each player can be measured for time intervals other than the entire match; ii) when appropriate filtering is applied to the aggregation process, centrality can be applied to the player's activity in attacking plays of certain characteristics (e.g. plays that lead to a shot at goal event).

2.3.2. Question 2: which players have an intermediary role?

Betweenness[5] is a widely used centrality metric that could provide an answer to the important question of which players play an intermediary role. Betweenness analyzes the *global structure* of the network, notably the fraction of shortest paths corresponding to each node, and it shows potential for accurately representing how much each player contributes as an intermediary between other players. However, this metric is typically based on the shortest path between any two nodes computed over the graph resulting from the aggregation of all the interactions in the match (17). In these conditions the metric is not directly supported by any fundamental concept of an attacking play and this is, in our view, a strong limitation. Indeed, in team ball sports the ball flow does not

---

$$Centrality_{degree}(v_i) = \text{degree}(v_i) = \sum_{j}^{N} a_{ij}$$

where i is the focal node, j represents all other nodes, N is the total number of nodes, and a is the adjacency matrix, in which cell a_{ij} is defined as 1 if node i is connected to node j, and 0 otherwise.

[5] *Betweenness centrality* expresses the degree in which one node lies on the shortest path between two other nodes;

$$Centrality_{betweenness}(v_i) = betweenness_i = \frac{g_{st}(i)}{g_{st}}$$

where g_{st} is the number of shortest paths between vertices s and t, and $g_{st}(i)$ is the number of those paths that pass through vertex i.

necessarily follow the shortest paths over the aggregation graph; instead players projections reveal mostly walks (with relevant levels of randomness) and not paths, similarly to most flows in other networks (34). In addition, the paths that lead to specific events (e.g. shot at goal) may have a more direct impact on team performance than the connectivity between players per se (19). Moreover, how such connectivity is directed in specific events can also affect team performance (12). Nevertheless, these limitations of betweeness can be overcome by counting the fraction of walks leading to a certain event in which the focal player is involved, rather than considering the shortest paths between players (12). Freeman and colleagues (16) proposed a metric based on the idea of maximum flow, *flow centrality*, whereby different paths can be used for the same purpose and which has been applied to basketball research by Fewell and colleagues (19). Newman's *random-walk betweenness* considers all paths between nodes including those that are not optimal, although more weight is given to the shortest paths (34). Bonacich (35) refers to *power centrality*, a metric based on the assumption that centrality is related to power and as such an individual's status is a function of the statuses of his/her connections.

2.3.3. Question 3: how central is a player?

Comparing players in terms of the number of passes achieved can reveal important individual characteristics of a player's performance and also the soccer team's style of play. *Closeness*[6] is a measure of centrality that considers the length of the shortest paths between the focal node and all the other nodes. We propose that this metric should be used as follows: passing path projections convey length (measured by the absolute frequency of passes) and duration of each play. The latter can be used directly to identify the team's style of play (see Passos et al. [1]). Moreover, by analyzing passing path projections (see Figure 3) one can capture the distance (also measured by the absolute frequency of passes) between the focus player and an event of interest, which can be utilized for example to identify which players contribute directly to shooting at goal or assisting other players. Alternatively, Noh and Rieger's (36) *random-walk centrality* metric describes the average speed at which messages are transmitted from

[6] *Closeness centrality* for each node, v_i, is the inverse sum of the shortest distance, $distance(i,j)$ to all other nodes, j, from the focal node, i, or how long the information takes to spread from a given node to others.

$$Closeness_{centrality}(v_i) = closenness_C(i) = \left[\sum_{j=1}^{N} distance(i,j)\right]^{-1}$$

one node to another in random walks (37). This is similar to the *closeness centrality* metric except that it considers the length of a random walk rather than the shortest path.

2.3.4. Question 4: how does each player contribute to the performance of others?

A player can contribute to the team's performance directly (e.g. goals scored) and indirectly by assisting the performance of team-mates (e.g. assistances/passes). *Eigenvector centrality*[7] (38) can be used to assess the contribution of each player to the team's performance. Similar to closeness, eigenvector centrality considers the global structure of the network (29) but assumes unrestricted walks, rather than paths, emanating from a node. Thus, this measure counts the number of walks of all lengths, weighted inversely by length, and as a result it can determine how each node affects all its neighbours at a given moment (29). Alternatively, Bonacich (35) argues that the concept of centrality should be more general due to its positive relationship with power (39). This way Bonacich (35) proposes a *power centrality* metric related to power and hence the individual or node's status is dependent on the status of its connections. The power is therefore attributed to the nodes/players in the negotiation of any single play with their team-mates, and it results from the players' efficacy to resolve previous moves. Moreover, this metric can be complemented by pre-defining weights for relevant events in the graph, for instance, ball loss with a negative weight and scoring a goal with the highest positive value. These weights can then be propagated to other nodes/players, similarly to the eigenvector centrality measure.

2.3.5. Question 5: are there "hot" nodes in the team?

During a play situation, players tend to search for the team-mate who can typically offer more solutions to the problems of the game (40). This preference could be measured by identifying those players who connect more often with the 'powerful' ones. Passos and

---

[7] *Eigenvector centrality*, takes into consideration not only how many connections a vertex has (i.e., its degree), but also the *degree* of the vertices that it is connecting to. Each vertex i is assigned a weight $x_i > 0$, which is defined to be proportional to the sum of the weights of all vertices that point to i: $x_i = \lambda^{-1} \sum_j A_{ij} x_j$ for some $\lambda > 0$, or in matrix form

$$Ax = \lambda x,$$

where A is the (asymmetric) adjacency matrix of the graph, whose elements are A_{ij}, and x is the vector whose elements are the x_i, and λ is a constant (the eigenvalue).

colleagues (13) suggested that identifying these *preferential attachments*[8] could lead to the 'decision-makers' of each team. Grund (14) has provided some innovations in this regard by assessing team ball flow with a traditional binary system (pass =1, no pass = 0) but also by considering centralities based on node strength and ties weight. The author focused on network structure (41-44) thus confirming that increased interaction intensity (density) leads to increased team performance (measured by the absolute frequency of goals scored), and increased centralization of interactions leads to decreased team performance. Given the intrinsic temporal nature of the graphs describing a match, we suggest that preferential attachments could be assessed by determining if those players with the highest node degree are more likely to be selected in the next attacking plays. By applying appropriate algorithms to the aggregation process, it is possible to obtain distinct metrics for different attacking play outcomes, and therefore to determine whether the preferential attachment process is related to those outcomes.

2.3.6. Question 6: are there players promoting clusters in the team?

Fewell and colleagues (19) assessed basketball team dynamics through degree centrality, clustering, entropy, and flow centrality, in order to uncover the play strategies of the 2010 National Basketball Association play-offs. In a group of nodes/players, it is possible to identify the players who are mutually highly connected and those who are less so. Such highly-connected groups are called *clusters*. The local *clustering coefficient*[9], cc_i, can be used to measure this connectivity property (property

---

[8] *Preferential attachments*, also known as cumulative advantage or '*rich-get-richer paradigm'*.

This property means that every new vertex probability (p_i) to connect the existing vertices is higher for those who have already a large number of connections (connectivity k_i). For example, in a given team sports with ball, when a player attracts more interactions from the game's beginning, his/her connectivity will increase at a higher rate when compared to his/her team-mates as the game is played (network grows). Therefore, starting with a small number (m_0) of players interacting at the beginning of the game, at every time step that a new player $m(\leq m_0)$ interacts with m different team-mates already active in the game, for preferential attachment, there is a probability $p_i(k_i) = \frac{k_i}{\sum_j k_j}$ that the new player i will interact with a certain team-mate, depending on the connectivity k_i of the latter.

[9] The *local clustering coefficient* (cc_i) for player i is defined by the proportion of actual edges/interactions (e_i) between the $n_i \geq 2$ common neighbors of a vertex/player i and the number of possible edges between them.

$$cc_i = \frac{2e_i}{n_i(n_i-1)}$$

The *local clustering coefficient* over the aggregate of all plays (Figure 4) takes the following values:

$$cc_{GK} = 0, cc_{LD} = 1, cc_{RD} = 1,\ cc_{MF} = \frac{2}{3}, cc_{CF} = 1$$

known as transitivity in social networks) by capturing the probability of cooperation between players as a function of their mutual acquaintances/interactions. It can also be said that a *triadic closure* is formed around the focal node, i.e. "if two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future" (45, 46). However, when looking for the most relevant players in the formation of clusters, applying the local clustering coefficient to graphs representing an entire match (Figure 1) restricts its usefulness, as the passes (i.e. links) between players considered in the cluster formation may have occurred in different attacking plays, and possibly with a long temporal gap between them. Such is the case for the clique RD-MF-CF in Figure 4, showing an aggregation projection of a play that culminated with a shot at the goal. Given the attacking play granularity of bipartite and temporal networks (see Figure 3), only clusters formed within the same attacking play can be identified with certainty (e.g. in Figure 3, only the clique RD-LD-MF is identified). Taking advantage of the temporal structuring provided by attacking players' we can define a play local clustering coefficient[10], $cc_{i,j}$, computed for each attacking play and an aggregation local clustering coefficient[11], cc_i^*, averaging the former over a set of M_k attacking plays (k_1 to k_M) that compose aggregation k. The aggregation interval is thus not limited to the complete match aggregation and can be defined by events from a higher layer (e.g., all the attacking plays until a goal is scored).

GK goalkeeper, LD/RD left/right defender, MF midfielder, CF center forward

[10] The j^{th}*play local clustering coefficient* (cc_i) for player i in the j^{th}attacking is defined in a similar manner to the local clustering coefficient but takes into account only the players' projection network formed in the j^{th}attacking play.

The 2^{nd} *play local clustering coefficient*, $cc_{i,2}$, (Figure 3) takes the following values:

$$cc_{GK,2} = 0,\ cc_{LD,2} = 1,\ cc_{RD,2} = \frac{1}{2}, cc_{MF,2} = 1, cc_{CF,2} = 0$$

GK goalkeeper, LD/RD left/right defender, MF midfielder, CF center forward

[11] The k *aggregation local clustering coefficient* ($cc_{i,k}^*$) for player i is defined by the average of the local cluster coefficients for player i over the M_k (k_1 to k_M) attacking plays that compose the k aggregation.

$$cc_{i,k}^* = \frac{1}{M_k} \sum_{j=k_1}^{k_M} cc_{i,j},$$

The *aggregate play local clustering coefficient,* for the k aggregate composed of attacking plays 1 and 2, has the following values for each of the players:

$$cc_{GKk}^* = 0,\ cc_{LD,k}^* = \frac{1}{2}(0+1) = \frac{1}{2},\ cc_{RD,k}^* = \frac{1}{2}\left(0+\frac{1}{2}\right) = \frac{1}{4}, cc_{MF,k}^* = \frac{1}{2}, cc_{CF,k}^* = 0$$

GK goalkeeper, LD/RD left/right defender, MF midfielder, CF center forward

2.4 What is still to be done? The dynamics of a network and multilayer networks.

Although the questions we have addressed thus far could be answered through *static networks*[12] or by looking at the flows between nodes, our next questions require the analysis of changes in the network structure. In typical team interactions representations (described in section 2.2), only the interactions (flows) between players are considered for building the network. There is therefore a superposition between the flows *on* the network and the definition *of* the network structure that must be taken into account when considering the applicability of SNA metrics. However, these relationships occur throughout a certain time span, and these important time changes must also be considered.

The proposal in section 2.2 of using bipartite networks to identify nodes representing players and technical actions in different layers can be extended hierarchically. Technical actions (level *n* events) can be linked in another bipartite network to level *n+1* events corresponding to higher-level concepts, as illustrated in Figure 5.

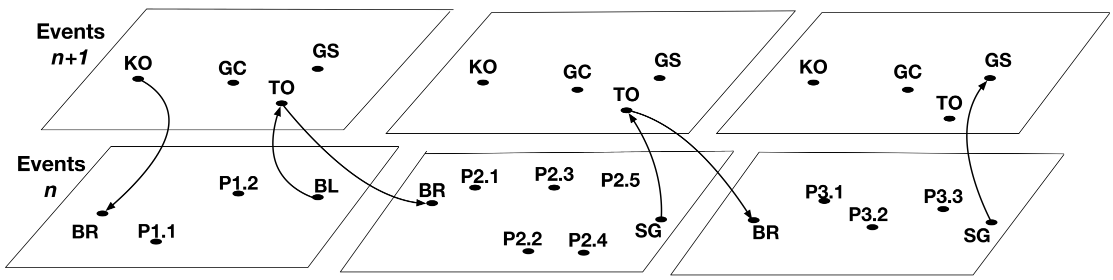


Figure 5: Hierarchical events represented through multilayer networks. KO kick off, TO turn over, GS goal scored, GC goal conceded, BR/BL ball recovery/loss, P pass, SG shot at goal. The first number represents the number of the play, the second number represents the order of the event. Similar metrics to the previously proposed can still be applied, although naturally corresponding to different concepts. For example, the concept of walk length applied to the *n* and *n*+1 level relationships reveals the number of attacking plays till a goal is scored.

2.4.1. Question 7: how does a player influence the team structure?

It is relevant to ask how player behaviors other than passing influence the overall network of interactions in the team. According to Barzel and Barabási (47), behavior

[12] We define as static network the static structure resulting from the aggregation over a time interval (e.g., the entire match) of all the observable edges (e.g., passes) within that interval.

prediction in a complex system requires a quantitative description of the system's structure and dynamics. The *dynamics of a network* considers various phenomena, including self-organization, that promote changes in the topology of the network (25), however, this metric has not yet been applied to sports sciences. The study of the dynamical interplay between the players's state and the topology of the network is recent and mostly theoretical (25, 48-51). Indeed, few theorizations have been corroborated by empirical results (25, 52). Interestingly, one of these studies (40) revealed an unexpected time dependence in network centrality (*dynamic centrality*) indicating that well-connected nodes can quickly become weakly-connected or even disconnected (25). Moreover, *dynamic centrality* expressed in adaptive networks (dynamic scale-free [DSF] networks) emerges from a reinforcement rule whereby each node considers only the importance or popularity of its neighbours (25). These surprising results reinforce the need for further studies, especially in sports settings.

Guillaume and Latapy (53) proposed another relevant approach to team sports performance showing how all complex networks may be described as bipartite structures, or alternatively via hypernetworks. The authors introduced a model that can be tested for any type of real-world complex network. Moreover, these bipartite networks can be used to represent relations that are not dyadic (i.e. they involve more than two actors). Non-dyadic relations among players, such as geographical proximity in the pitch, can describe other dimensions of the players' actions during the game and these descriptions can be used to understand the dynamics *of* the network of passes, in particular, by explaining why certain spatial team configurations lead to specific pass paths.

2.4.2. Question 8: how does the adversary team constrain the team's interactions and structure?

The manner and extent to which the opposing team constrains a team's interactions and structure is a much more complex question, as it considers the influence of the adversary team (individuals and structure) on the interactions between team individuals as well as on the team's structure. Some of the PA existing studies focus on the interplay between attackers and defenders and are therefore based on dyadic interactions between players, sub-units or teams. Typically, such studies associate these interactions with the players' spatial organization, which is computed from the surface area in so-called centroids (23, 54, 55) or Voronoi diagrams[13] (56). However, in network theory, *spatial networks* represent the nodes and edges based on their interactions in an Euclidean space. Using these metrics, research on social networks and space has identified ordering principles such as *homophily states* (57-59) and *focus constraint* (59, 60). Notably, while *homophily* depends on non-structural features such as connections fostered by status or interests (e.g. dyadic attacker-defender interactions in team ball sports), *focus constraints* are dependent on geographical proximity, enabling face-to-face interactions (59).

3. Conclusions

In this article, we reviewed how PA emerged as a sub-discipline of sports sciences by building on notational analysis and biomechanics approaches and with further contributions from DST. Additionally, we discussed what new directions, tools and potential methods network theory and complex networks can further contribute to PA. Early studies with network methodologies in team ball sports mainly considered the dynamics on networks through ball flow, which represent the interactions between players of the same team, or actions to score.

However, these studies did not consider the dynamics of networks, assuming teams to be static structures, whereby players retain the same performance level throughout the entire game, independent of the constraints imposed by team adversaries and the players' positioning. Moreover, the players' skills and technical actions as well as the evolution of the interactions between players over time were also not considered.

Finally, this static network structure is obtained via the aggregation of all the interactions (passes) that occurred in an entire match. This may lead not only to concealment of important concepts such as attacking play but also to metrics that may be misleading. A notable example is metrics based on `shortest paths´ over the aggregated network, that occur in two scenarios: i) pass interactions form a walk and not necessarily a path; ii) interactions between players in a match do not follow this principle but may be better described by geographic networks and random walks (34).

We propose that temporal and bipartite networks could be an alternative approach for representing the interactions between players during a game. Using the flexible time

[13] Voronoi diagrams are geometric constructions that represent the nearest geographical region of a player, a sub-set of a team, or even a team.

structure of temporal networks it is possible to capture the sequence of passes in an attacking play, which is one of the main concepts of team collective behaviour. We have highlighted how temporal bipartite network representation empowers existing metrics for capturing sports fundamental concepts (e.g. style of play) with greater adequacy. Moreover, we suggest that methods combining spatial and hypernetworks (61) with temporal networks represent a promising direction for future research, as they allow the analysis of dynamics of the network. These complex networks could integrate concepts such as how time changes the structure of the network, as well as the players' technical resources and their positioning relative to the position of other players (team-mates or adversaries).

Compliance with Ethical Standards

Funding

This work was partly supported by the Fundação para a Ciência e Tecnologia, under Grant UID/DTP/UI447/2013 to CIPER – Centro Interdisciplinar para o Estudo da Performance Humana (unit 447).

Conflicts of Interest

João Ramos, Rui J. Lopes and Duarte Araújo declare that they have no conflicts of interest relevant to the content of this review.

References

1. Passos P, Araújo D, Volossovitch A. Performance analysis in team sports. London: Routledge. Taylor & Francis Group.; 2017.
2. Glazier PS. Game, set and match? Substantive issues and future directions in performance analysis. Sports Med. 2010;40(8):625-34.
3. Hughes M, Franks IM. The essentials of performance analysis : an introduction. London ; New York: Routledge; 2008. xxxii, 312 p. p.
4. Frencken W. Soccer tactics: dynamics of small-sided games and full-sized matches. Groningen: University of Groningen - Faculty of Medical Sciences; 2012.
5. Vilar L, Araújo D, Davids K, et al. The role of ecological dynamics in analysing performance in team sports. Sports Med. 2012;42(1):1-10.
6. Dutt-Mazumder A, Button C, Robins A, et al. Neural network modelling and dynamical system theory: are they relevant to study the governing dynamics of Association Football players? Sports Med. 2011. 41: 1003-17. doi:10.2165/11593950-000000000-00000.
7. Bartlett R, Button C, Robins M, et al. Analysing team coordination patterns from player movement trajectories in soccer: methodological considerations. Int J Perform Anal Sport. 2012;12(2):398-424.
8. Davids K, Araújo D, Shuttleworth R. Applications of dynamical systems theory to football. Science and football V. 2005:537-50.
9. Araújo D, Davids K. Team synergies in sport: Theory and measures. Front Psychol. 2016;7.
10. Chinellato DD, de Aguiar MA, Epstein IR, et al. Dynamical response of networks under external perturbations: Exact results. arXiv preprint arXiv:07054607. 2007.
11. Silva P, Chung D, Carvalho T, et al. Practice effects on intra-team synergies in football teams. Hum Mov Sci. 2016;46:39-51.
12. Duch J, Waitzman JS, Amaral LAN. Quantifying the performance of individual players in a team activity. PLoS ONE. 2010;5(6):e10937.
13. Passos P, Davids K, Araújo D, et al. Networks as a novel tool for studying team ball sports as complex social systems. J Sci Med Sport. 2011;14(2):170-6.
14. Grund TU. Network structure and team performance: The case of English Premier League soccer teams. Soc Networks. 2012;34(4):682-90.
15. Travassos B, Bourbousson J, Esteves PT, et al. Adaptive behaviours of attacking futsal teams to opposition defensive formations. Hum Mov Sci. 2016;47:98-105.
16. Freeman LC. Visualizing social networks. J Soc Struct. 2000;1(1):4.
17. Rocha L. Exploring patterns of empirical networks. UmeÃ¥: UmeÃ¥ University, Department of Physics; 2011. 66 p.
18. Moody J, McFarland D, Bender-Demoll S. Dynamic network visualization. AJS. 2005;110(4):1206-41.
19. Fewell JH, Armbruster D, Ingraham J, et al. Basketball teams as strategic networks. PLoS ONE. 2012;7(11):e47445.
20. Newman MEJ. Networks : an introduction. Oxford ; New York: Oxford University Press; 2010. xi, 772 p. p.
21. Brandes U, Freeman L, Wagner D. Social Networks. In: (ed.) RT, editor. Handbook of Graph Drawing and Visualization. CRC Press2012.

22. Eccles DW, Tenenbaum G. Why an expert team is more than a team of experts: A cognitive conceptualization of team coordination and communication in sport. J Sport Exerc Psychol. 2004;26:542-60.
23. Duarte R, Araújo D, Freire L, et al. Intra-and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. Hum Mov Sci. 2012;31(6):1639-51.
24. Bender-deMoll S, McFarland DA. The art and science of dynamic network visualization. J Soc Struct. 2006;7(2):1-38.
25. Hill SA, Braha D. Dynamic model of time-dependent complex networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2010;82(4):046105.
26. Araújo D, Lopes RJ, Ramos J, et al. A final à lupa: Alemanha pressionante, Argentina recuada. Lisbon. Portugal: Jornal Público Online; 2014 [Available from: https://http://www.publico.pt/desporto/noticia/a-final-a-lupa-pressao-sustentada-da-alemanha-venceu-uma-argentina-recuada-e-a-procura-do-contraataque-1662863.
27. Lucchesi M. Attacking Soccer: A Tactical Analysis. 2001. Spring City, PA: Reedswain Publishing.
28. Rocha L, Masuda N. Random walk centrality for temporal networks. New J Phys. 2014;16(6):063023.
29. Borgatti SP. Centrality and network flow. Soc Networks. 2005;27(1):55-71.
30. Pena JL, Touchette H. A network theory analysis of football strategies. arXiv preprint arXiv:12066904. 2012.
31. Clemente FM, Martins FML, Kalamaras D, et al. General network analysis of national soccer teams in FIFA World Cup 2014. Int J Perform Anal Sport. 2015;15(1):80-96.
32. Clemente FM, Couceiro MS, Martins FML, et al. Using network metrics in soccer: A macro-analysis. J Hum Kinet. 2015;45(1):123-34.
33. Correia V, Araújo D, Duarte R, et al. Changes in practice task constraints shape decision-making behaviours of team games players. J Sci Med Sport. 2012;15(3):244-9.
34. Newman MEJ. A measure of betweenness centrality based on random walks. Soc Networks. 2005;27(1):39-54.
35. Bonacich P. Power and centrality: A family of measures. AJS. 1987:1170-82.
36. Noh JD, Rieger H. Random walks on complex networks. Phys Rev Lett. 2004;92(11):118701.
37. Stephenson K, Zelen M. Rethinking centrality: Methods and examples. Soc Networks. 1989;11(1):1-37.
38. Bonacich P. Factoring and weighting approaches to status scores and clique identification. J Math Sociol. 1972;2(1):113-20.
39. Cook KS, Emerson RM, Gillmore MR, et al. The distribution of power in exchange networks: Theory and experimental results. AJS. 1983:275-305.
40. Braha D, Bar-Yam Y. From centrality to temporary fame: Dynamic centrality in complex networks: Research Articles. Complex. 2006;12(2):59-63.
41. Reagans R, Zuckerman EW. Networks, diversity, and productivity: the social capital of corporate R&D teams. Organization Science. 2001;12(4):502-17.
42. Sparrowe RT, Liden RC, Wayne SJ, et al. Social networks and the performance of individuals and groups. Acad Manage J. 2001;44(2):316-25.
43. Balkundi P, Kilduff M. The ties that lead: A social network approach to leadership. Leadersh Q. 2006;17(4):419-39.
44. Cross R, Cummings JN. Tie and network correlates of individual performance in knowledge-intensive work. Acad Manage J. 2004;47(6):928-37.

45. Rapoport A. Spread of information through a population with socio-structural bias: I. Assumption of transitivity. Bull Math Biophys. 1953;15(4):523-33.
46. Easley D, Kleinberg JY. : Networks, crowds, and markets: reasoning about a highly connected world. Cambridge University Press. ISBN: 0521195330, 9780521195331. 2010.
47. Barzel B, Barabási A-L. Universality in network dynamics. Nat Phys. 2013;9(10):673-81.
48. Holme P, Ghoshal G. Dynamics of networking agents competing for high centrality and low degree. Phys Rev Lett. 2006;96(9):098701.
49. Pacheco JM, Traulsen A, Nowak MA. Coevolution of strategy and structure in complex networks with dynamical linking. Phys Rev Lett. 2006;97(25):258103.
50. Van Segbroeck S, Santos FC, Pacheco JM. Adaptive contact networks change effective disease infectiousness and dynamics. PLoS Comput Biol. 2010;6(8):e1000895.
51. Shaw LB, Schwartz IB. Fluctuating epidemics on adaptive networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2008;77(6):066101.
52. Barabasi A-L. The origin of bursts and heavy tails in human dynamics. Nature. 2005;435(7039):207-11.
53. Guillaume J-L, Latapy M. Bipartite graphs as models of complex networks. Physica A. 2006;371(2):795-813.
54. Lames M, Erdmann J, Walter, F. Oscillations in football – order and disorder in spatial interactions between the two teams . Int J Sport Psychol. 2010;41:85-6.
55. Frencken W, Lemmink K, Delleman N, et al. Oscillations of centroid position and surface area of soccer teams in small-sided games. Eur J Sport Sci. 2011;11(4):215-23.
56. Fonseca S, Milho J, Travassos B, et al. Measuring spatial interaction behavior in team sports using superimposed Voronoi diagrams. Int J Perform Anal Sport. 2013;13(1):179-89.
57. Kossinets G, Watts DJ. Empirical analysis of an evolving social network. Science. 2006;311(5757):88-90.
58. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu Rev Sociol. 2001;27(1):415-44.
59. Expert P, Evans TS, Blondel VD, et al. Uncovering space-independent communities in spatial networks. Proc Natl Acad Sci USA. 2010;108(19):7663–7668. doi:10.1073/pnas.1018962108.
60. Crandall DJ, Backstrom L, Cosley D, et al. Inferring social ties from geographic coincidences. Proc Natl Acad Sci USA. 2010;107 (52) 22436-22441; doi:10.1073/pnas.1006155107.
61. Johnson JH, Iravani P. The multilevel hypernetwork dynamics of complex systems of robot soccer agents. ACM Transactions on Autonomous and Adaptive Systems (TAAS). 2007;2(2):5.

3.2 Hypernetworks Reveal Compound Variables That Capture Cooperative and Competitive Interactions in a Soccer Match

3.2.1 Context and summary

This study aims to overcome SNA limitations by using hypernetworks to describe illustrative cases of team behavior dynamics at various other levels of analyses. Hypernetworks simultaneously access cooperative and competitive interactions between teammates and opponents across space and time during a match. Moreover, hypernetworks are not limited to dyadic relations, which are typically represented by edges in other types of networks. In a hypernetwork, *n-ary* relations (with n>2) and their properties are represented with hyperedges connecting more than two players simultaneously (the so-called *simplex* – plural, *simplices*). Simplices can capture the interactions of sets of players that may include an arbitrary number of teammates and opponents. In this qualitative study, we first used the mathematical formalisms of hypernetworks to represent a multilevel team behavior dynamics, including micro (interactions between players), meso (dynamics of a given critical event, e.g., an attack interaction) and macro (interactions between sets of players) levels. Second, we investigated different features that could potentially explain the occurrence of critical events, such as aggregation or disaggregation of simplices relative to goal proximity. Finally, we applied hypernetworks analysis to soccer games from the English premier league (season 2010-2011) by using two-dimensional player displacement coordinates obtained with a multiple-camera match analysis system provided by STATS (formerly Prozone).

We have extended the approach by Johnson & Iravani (2007) by introducing compound variables, e.g. local dominance, which capture the structure and dynamics of cooperative and competitive interactions.

The aim of this study was therefore to operationalize a method addressing different levels of hypernetworks on soccer matches.

The results show that:

i) At micro level the most frequently occurring simplices configuration by decreasing order of frequency are: 1vs.1, 2vs.1 and 1vs.2, 2vs.2, and finally, 3vs.1 and 1vs.3. However, these simplices show differences in their distribution on the pitch, and this is particularly evident for unbalanced simplices such as 2vs.1, 1vs.2, 3vs.1 and 1vs.3. These differential distributions are consistent with the match result (wins vs. losses) and the opponent team's strength;

ii) At meso level, the dynamics of simplices transformations near the goal depends on significant changes in the players' speed and direction, to improve their positioning to score or to unbalance the situation.

iii) At macro level, simplices are connected to one another, forming "simplices of simplices" including the goalkeeper and the goal.

These results may significantly contribute to improve training and playing strategies and therefore to validate qualitatively that hypernetworks and related compound variables can capture and be used in the analysis of the cooperative and competitive interactions between players and sets of players' in soccer matches.

frontiers in Psychology

METHODS
published: 28 August 2017
doi: 10.3389/fpsyg.2017.01379

3.2.2 Paper *verbatim* copy

Hypernetworks Reveal Compound Variables That Capture Cooperative and Competitive Interactions in a Soccer Match

João Ramos[1,2], Rui J. Lopes[1,3], Pedro Marques[4] and Duarte Araújo[5]*

[1] ISCTE-Instituto Universitário de Lisboa, Lisbon, Portugal, [2] Universidade Europeia, Laureate International Universities, Lisboa, Portugal, [3] Instituto de Telecomunicações, Lisbon, Portugal, [4] Football Performance, City Football Services, Manchester, United Kingdom, [5] Interdisciplinary Centre for the Study of Human Performance (CIPER), Faculdade de Motricidade Humana, Universidade de Lisboa, Lisbon, Portugal

The combination of sports sciences theorization and social networks analysis (SNA) has offered useful new insights for addressing team behavior. However, SNA typically represents the dynamics of team behavior during a match in dyadic interactions and in a single cumulative snapshot. This study aims to overcome these limitations by using hypernetworks to describe illustrative cases of team behavior dynamics at various other levels of analyses. Hypernetworks simultaneously access cooperative and competitive interactions between teammates and opponents across space and time during a match. Moreover, hypernetworks are not limited to dyadic relations, which are typically represented by edges in other types of networks. In a hypernetwork, *n-ary* relations (with $n > 2$) and their properties are represented with hyperedges connecting more than two players simultaneously (the so-called *simplex*—plural, *simplices*). Simplices can capture the interactions of sets of players that may include an arbitrary number of teammates and opponents. In this qualitative study, we first used the mathematical formalisms of hypernetworks to represent a multilevel team behavior dynamics, including micro (interactions between players), meso (dynamics of a given critical event, e.g., an attack interaction), and macro (interactions between sets of players) levels. Second, we investigated different features that could potentially explain the occurrence of critical events, such as, aggregation or disaggregation of simplices relative to goal proximity. Finally, we applied hypernetworks analysis to soccer games from the English premier league (season 2010–2011) by using two-dimensional player displacement coordinates obtained with a multiple-camera match analysis system provided by STATS (formerly Prozone). Our results show that (i) at micro level the most frequently occurring simplices configuration is 1vs.1 (one attacker vs. one defender); (ii) at meso level, the dynamics of simplices transformations near the goal depends on significant changes in the players' speed and direction; (iii) at macro level, simplices are connected to one another, forming "simplices of simplices" including the goalkeeper and the goal. These results validate qualitatively that hypernetworks and related compound variables can capture and be used in the analysis of the cooperative and competitive interactions between players and sets of players in soccer matches.

Keywords: network theory, hypernetworks, network dynamics, performance analysis, soccer

OPEN ACCESS

Edited by:
*Nancy J. Cooke,
Arizona State University, United States*

Reviewed by:
*Jamie Gorman,
Georgia Institute of Technology, United States
Mustafa Demir,
Arizona State University, United States
Aaron Likens,
Arizona State University, United States*

****Correspondence:***
*João Ramos
jpdrs@iscte-iul.pt*

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Received: *20 April 2017*
Accepted: *31 July 2017*
Published: *28 August 2017*

Citation:
Ramos J, Lopes RJ, Marques P and Araújo D (2017) Hypernetworks Reveal Compound Variables That Capture Cooperative and Competitive Interactions in a Soccer Match. Front. Psychol. 8:1379. doi: 10.3389/fpsyg.2017.01379

INTRODUCTION

Coaches, players, and scientists have long tried to understand team behavior dynamics during a game, aiming to develop interventions and training plans that may increase team performance (Araújo and Davids, 2016; Passos et al., 2017). Broadly speaking, research in performance analysis in team sports searches for variables describing game dynamics that are: (i) useful and accessible to coaches and athletes; (ii) obtained automatically or semi-automatically from game observation; and (iii) related to team outputs, such as, match results. For finding such variables it is necessary to capture the multi-leveled dynamics emerging from differential interactions between many heterogeneous parts (e.g., players), while considering potential adaptations to changing environments. In this way, teams and athletes can be seen as co-evolving subsystems that self-organize into new structures and behaviors (Johnson, 2013), i.e., they form team synergies (Araújo and Davids, 2016). Such team synergies emerge from physical and informational constraints (Schmidt et al., 1998, 2011). Importantly players are perceptually linked mainly by informational constraints, since physical links among them are very rare (e.g., when forming a wall of players; Riley et al., 2011). Several studies have analyzed the coupling among performers based on interpersonal distance measures (Passos et al., 2011; Fonseca et al., 2013; Rio et al., 2014), with a higher emphasis on the distance between a player and the immediate opponent (e.g., Headrick et al., 2012). In the present study, we extend this player-immediate opponent distance to the closest player (opponent or not).

These interactions, based on informational and physical constraints have been studied by network theorical approaches, like social network analysis (SNA). SNA is a powerful tool to capture and study interpersonal relations in team sports (Araújo and Davids, 2016); however, this method can only be used for representing binary (*2-ary*) relations (Johnson, 2006; Criado et al., 2010; Boccaletti et al., 2014). The most common graphical representations of SNA depict players as nodes in fixed positions in the pitch (the field of the match), with edges between them representing the cumulative "ball flux," i.e., ball passes, over time (Duch et al., 2010; Fewell et al., 2012; Grund, 2012; Clemente et al., 2015; Araújo and Davids, 2016; Travassos et al., 2016). This is a fundamental limitation of typical SNA in sport context, as it restricts its application to the attacking phase of team dynamics. Typically, all other relevant types of interactions, either cooperative or competitive, are not considered. In this study, we investigate how cooperative (e.g., between players of the same team in order to create a scoring opportunity) and competitive interactions (e.g., between players of different teams competing for ball possession) may be captured and analyzed via multilevel hypernetworks. On the one hand, according to Boccaletti et al. (2014), multilevel networks constitutes the new frontier in many areas of science since it describes systems that are interconnected through different categories of connections (e.g. relationship: teammate vs. opponent; activity: increasing vs. diminishing interpersonal distance; category: attacker vs. midfielder) that can be represented in multiple layers, including networks of networks (e.g., interactions between teams). On the other hand, in a hypernetwork, a hyperedge can connect more than two nodes, thus directly representing *n-ary* interactions occurring among small sets of nodes, $\langle p_i, \ldots, p_j \rangle$ (Johnson, 2006, 2008, 2013, 2016; Criado et al., 2010; Boccaletti et al., 2014). This generalization provided by hypernetworks enables the representation of cooperative and competitive interactions that occur during the game and that involve an arbitrary number of players (teammates or opponents).

In the present study, we have extended the approach by Johnson and Iravani (2007) by introducing compound variables, e.g., local dominance, which capture the structure and dynamics of cooperative and competitive interactions in the following ways:

i. By considering the domain specificity of soccer matches to tag the sets of players formed (e.g., 2 vs. 1 corresponds to a set with two attackers and one defender) as these tags describe local dominance (Duarte et al., 2012);
ii. By including the spatiotemporal occurrence of the different sets of players by counting their frequency and location;
iii. By analyzing and relating the dynamics of the sets with players velocity in specific events (goal scoring opportunities);
iv. By studying, for the same events of interest, the formation and dynamics of higher level simplices; notably, the relations between simplices of simplices.

The present approach is applied to a set of matches in order to investigate how the proposed compound variables can be useful on characterizing the behavior of players and teams at different levels and the relationships between these levels and match context, e.g., team local dominance and current match result.

As a first step in this approach, it is necessary, at each level of analysis, to identify the meaningful relations for the match dynamics, and represent them using different criteria for selecting the players in each set (i.e., connected by a hyperedge; Johnson, 2008, 2016). According to Passos and colleagues the analysis of the interpersonal distances is adequate for complex systems modeling (Passos et al., 2011). As we are interested in cooperative and competitive behavior in the pitch, geographical proximity between players (Headrick et al., 2012) can capture whether an interaction between players exists or not (e.g., functional couplings). Also, in the investigation of the relation between higher (macro) level of analysis and players' individual actions (micro), it is important to consider the velocity of each player, as well as the velocity of the set of players, represented by the set's geometric center and obtained through the computation of each players' velocity. For example if such set is expected to maintain its structure or if it is about to split when a player's velocity vector is moving away from the other players. Operationally, we have defined that a player does interact with his closest player; this interaction is cooperative when that closest player is a teammate, and competitive when it is an opponent. Thus, time and space are highlighted in the present approach using hypernetworks because it uses geographical proximity criteria, and also because it captures temporal changes, by considering the players' geographical positions over time

(t_1, $t_{2, \ldots}$, t_n). The compound variables adopted in this study reflect and capture this space and temporal features, e.g., local dominance and the dynamics, i.e., changes on, players' sets.

In **Figure 1**, we show an example of a set of nodes identified at Level N: two attacking players (a_1 and a_2,), a defender (d_1), a goalkeeper (d_0), and a goal (G_a). These nodes are connected by two hyperedges at Level $N + 1$, corresponding to sets $\langle a_1, a_2, d_1 \rangle$ and $\langle d_0, G_a \rangle$ in one time frame, and $\langle a_1, d_1 \rangle$ and $\langle a_2, d_0, G_a \rangle$ on the next.

For a more complete description of the system's dynamics, each tuple identified in the hypernetwork can be extended by an element, R, that describes the relationships in the set (Johnson, 2013). Each of these extended sets is called a *simplex* (Johnson and Iravani, 2007; Johnson, 2013). For example, R is the path to understand why the sets $\langle a_1, a_2, d_1 \rangle$ and $\langle d_0, G_a \rangle$ on one frame lead to the sets and $\langle a_1, d_1 \rangle$ and $\langle a_2, d_0, G_a \rangle$ on the next. When a player observes the game searching for the best action possibilities offered by the other players' positioning, the entire configuration of team-mates and opponents has to be perceived. Such sets of players, either in 1vs.1, 2vs.1, or 2vs.2, or any other set, may be related to one another, regarding the players' general configuration. Thus, when one player decides to move, the entire configuration is affected. Johnson and Iravani (2007) propose naming the "2 attackers vs. 1 defender" structure, the *defenders' dilemma*, since the defenders can opt to tackle the ball or intercept the pass between attackers. In a similar situation involving the goalkeeper, the *goalkeepers' dilemma*, the options are moving to the right or left of the goal, or moving toward the attacker leaving the goal behind. The goal can therefore be considered as a constraint that attracts the opponents and instigates the defenders to position as if it were an opponent. For this reason, we have included goals in the definition of simplices, because they show similarities to an "attacking player" (e.g., in the goalkeepers dilemma).

In this study, we propose several compound variables to describe the players' cooperative and competitive behavior dynamics during a soccer match. The simplest of these variables depicts the dominant interactions in each set, and is expressed by two values representing the number of attacking and defending players, for example, 2 vs. 1 corresponds to a set with two attackers and one defender. In **Figure 1**, the two dominant relationships are $R_1 = (2\ vs.1)$ and $R_2 = (0\ vs.1)$, and the corresponding simplices are $\sigma_1 = \langle a_1, a_2, d_1; (2\ vs.1) \rangle$ and $\sigma_2 = \langle d_0, G_a; (0\ vs.1) \rangle$. The behavior of a team during a match can then be described by other compound variables that characterize the relative frequencies of the aforementioned relationships. For example, the minimal structure (simplex) of players' interactions occurring more frequently in a match can be assessed.

At higher complexity levels, the hypernetwork can represent the interactions between related simplices, or simplices of simplices (see **Figure 1**, *Level N + 3;* Johnson, 2006, 2013; Johnson and Iravani, 2007). In what regards the study of dynamics: less dynamic structures (e.g., number of players, players' roles, etc.) are called *backcloth*, and higher rate changes

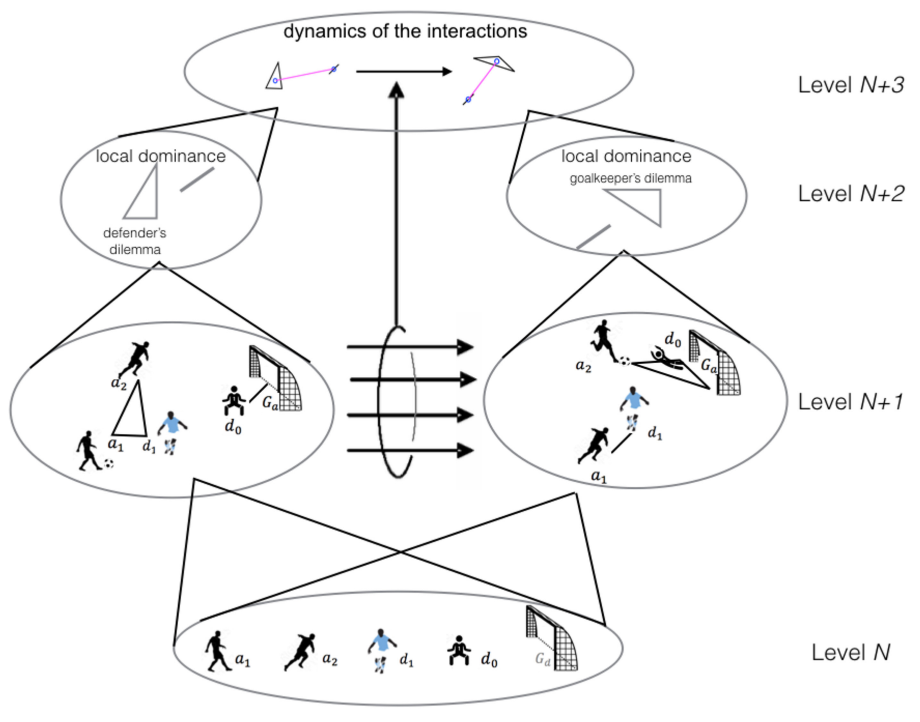


FIGURE 1 | Multilevel hypernetwork representation (from bottom to top). Each level corresponds to a different abstraction level (Level N, players in the pitch; Level $N + 1$, proximity-based simplices; Level $N + 2$, local dominance relation; Level $N + 3$, dynamic analysis via simplices of simplices). Also represented, the displacement in a soccer game of 2 sequential time frames (from the left to the right hand side) (Adapted from Johnson and Iravani, 2007).

(e.g., players positioning in relation to opponents, teammates and the goal or the ball) are called *traffic* (Johnson, 2013) and represent dynamics within the backcloth. Thus, one important feature of hypernetwork analysis in the sports context is the representation of players' *moves,* across time and space, and between structured sets (i.e., from one simplex to another). As shown in **Figure 1**, this multilevel approach allowed us to capture the number of players and their moves and the players in the match-day squad (*Level N*), the coordinated sets of players along the match (*Level N + 1*), the local advantage of one team over the other (e.g., numerical dominance; *Level N + 2*), and the relationship between the sets (*Level N + 3*). Moreover, by using this approach different compound variables, e.g., local dominance, may explain distinctive aspects of the competitive and cooperative behavior of players and teams.

In this study we put forward the hypothesis that hypernetworks and compound variables over these hypernetworks can capture relevant features of soccer team dynamics during a match. We validate qualitatively this hypothesis by applying the proposed method to a set of matches of a focal team within different contexts and by analysis the results thus obtained. The aim of this study was therefore to operationalize a method addressing different levels of hypernetworks on soccer matches and by providing a study case for tackling the following questions:

i. At *Level N*: Has the backcloth (players) changed during the match, as expressed by events such as, substitutions, sent-offs and injuries? Typical notational analyses answer this question directly.
ii. At *Level N + 1*: What are the most frequently occurring simplices in soccer matches? A histogram with the relative frequencies of occurrence of every type of simplices (e.g., 1vs.1, 2vs.1...) can be computed.
iii. At *Level N + 1*: Are there any differences in simplices' structure and occurrence between home or away matches for Team A? A heat map (2D spatial frequency map) for each of the relationships can be computed to show their location in the pitch.
iv. At *Level N + 1*: Are there any changes in simplices structure and field position as the match score changes? Instead of considering the entire match, the heat maps can address specific periods of the match. These periods are bounded by relevant match events, e.g., a goal being scored.
v. At *Level N + 2*: What are the dynamics of the simplices' interactions near the goal, immediately before the score changed? Instead of examining the results for the entire match, or for given periods, it is possible to perform a frame-by-frame analysis to assess which simplices formed and how they changed, and also to identify the players who contributed to those changes.
vi. At *Level N + 3*: Is there any interaction between simplices leading to the emergence of new team configurations that, in turn, can lead to scoring a goal? To answer this question, it is necessary to evaluate how the different simplices relate to one another, how they aggregate into higher-level simplices, and how they recombine into different simplices.

METHODS AND MATERIALS

Five matches were analyzed from a pool of 11 matches of the English Premier League season 2010–2011 provided by STATS (formally Prozone). This data set was selected because it contained no errors, such as, missing or duplicated positioning data, and because the *backcloths* were equivalent (i.e., there were no differences between teams regarding the number of players due to sent-offs or injuries without substitutions). Participants included all the players in the field from Team A (our focal team), and the players from five teams playing against team A (teams B, C, D, E, and F). The matches included three home matches, against teams B, C, and D, and two away matches, against teams E and F. The players' substitutions were considered but not analyzed in detail in this study (i.e., data for both initial squad and substitutes are used but the implications of substitutions in the backcloth are not taken into consideration).

Matches and their score were: Team A vs. Team B (1–0); Team A vs. Team C (1–0); Team A vs. Team D (1–0); Team E vs. Team A (2–1) and Team F vs. Team A (0–0). The details for each match are presented in **Table 1**.

For each match, raw data consisted of two-dimensional player displacement coordinates provided by STATS. These data were obtained by a multiple-camera match analysis system whereby the movements of the 22 players during the match were recorded with eight cameras positioned at the top of the stadium. The frames were processed at 10 Hz through an automated system that synchronized the video files. The effective playing area was 80 m wide and 120 m long, including the out-of-bound locations such as, set-plays. A computer procedure for computing the simplices' hyperedges set with the proximity criterion was implemented using GNU Octave version 4.2.0 and applied to each frame. This criterion has the advantage of being non-parametric; the corresponding pseudo-code for this algorithm is provided in **Figure A1**.

Each simplex was represented graphically by the convex hull computation (the minimum convex area containing all players in the simplex) and included the velocity of each player (vector velocity considering the instant *t-1* and *t*), as well as the velocity of the geometric center of the simplices.

To represent the field positioning of the different types of simplices, we used heat maps for the frequency of simplices occurrence. This type of graphical representation allowed us to capture the most frequent type of simplices for each time period, as well as their geographical position in the field.

TABLE 1 | Matches' details indicating the result and changes in the team structure due to sent-offs, substitutions, or injuries (without substitution).

Matches	A vs. B	A vs. C	A vs. D	E vs. A	F vs. A
Results	1–0	1–0	1–0	2–1	0–0
Substitutions	3–3	3–3	3–3	3–3	2–2
Sent-offs	0–0	0–0	0–0	0–0	1–1
Injuries (without substitution)	0–0	0–0	0–0	0–0	0–0

For analyzing specific time points, we represented simplices (*Level N + 2*, **Figures 5**, **6**) with two different colors: for players in team A, vertices are in **red**, for players in team B, vertices are in **green**. For the higher-level simplices in level *N + 3*, **Figure 6**, the blue **o** symbol represents the geometric center of the simplices. Such representation facilitates the simultaneous identification of players in both teams and the type of simplices in level *N + 3*. Moreover, we also represented the proportion (local dominance or balance) of each type of simplices in level *N + 2*, as well as the type of relation that exists between the simplices, or simplices of simplices in any instant of time at level *N + 3*. The velocity of the simplices and players were also included, thus allowing for the evaluation of simplices consistency, for example, transformations such as, when a player entered or moved away from a given simplice, or when all players moved simultaneously to the same position, could be detected.

RESULTS

Our results revealed how the matches' hypernetworks are characterized from *Level N* to *Level N + 3*.

We analyzed the structure at *Level N* of the five matches. As expected, we found 11 players in each team, with some players being substituted but with no sent-offs (with the exception of match F vs. A) or injuries occurring after there were no substitutions left (hence the total number of players remained constant). At this level of analysis, individual player statistics and heat maps of their positioning during the match are usually performed. However, as this type of performance analysis is widespread in sport (for a review see Passos et al., 2017), and given that the focus of this paper is on team behavior, we do not present such results here.

We computed the relative frequencies of the simplices structures at Level *N + 1* for players in both teams (**Figure 2**). The most frequently occurring simplices structures in the 5 matches: 1vs.1; 2vs.1; 1vs.2; 2vs.2; 3vs.1; 1vs.3. These results reveal that the most frequently occurring simplices structures are similar in every match. Around 25% of the simplex structures corresponded to 1vs.1, independently of the type of match (home or away) or its final result. The second most frequently occurring simplices structures were 2vs.1 and 1vs.2 (around 10%), followed by 2vs.2 (around 6%), and finally by 3vs.1 and 1vs.3 (around 3%). Among other simplices structures, we could also often find interactions between the goalkeeper and the goal, as identified in 0vs.1 or 1vs.0 structures (around 11%). However, these simplices structures do not reveal a social interaction (i.e., cooperation or competition) and are therefore not compared to other structures.

By computing the frequencies for the "local dominance tag" compound variable it is possible to investigate for each game the most frequent cooperation and competition interactions sets.

Level N + 1 describes the geographical distribution in the pitch of the most frequently occurring simplices structures, as shown in *heat maps* (**Figure 3**).

Figure 3 shows that although 1vs.1 is the most frequently occurring simplex tag in every match, the location in the pitch where it can more often be found varies between matches. Simplices, 2vs.1, indicating simultaneous cooperation and competition, occurs mostly in the mid-field, and simplices 1vs.2 occurs mostly in the opponent side of the field.

By identifying the relevant events in a match, such as, changes in the score, at *Level N + 1* we can capture changes in collective behavior across time. **Figure 4** shows the results of this analysis in heat maps corresponding to different sections of the E vs. A match (final result 2–1). For example, these heat maps reveal that

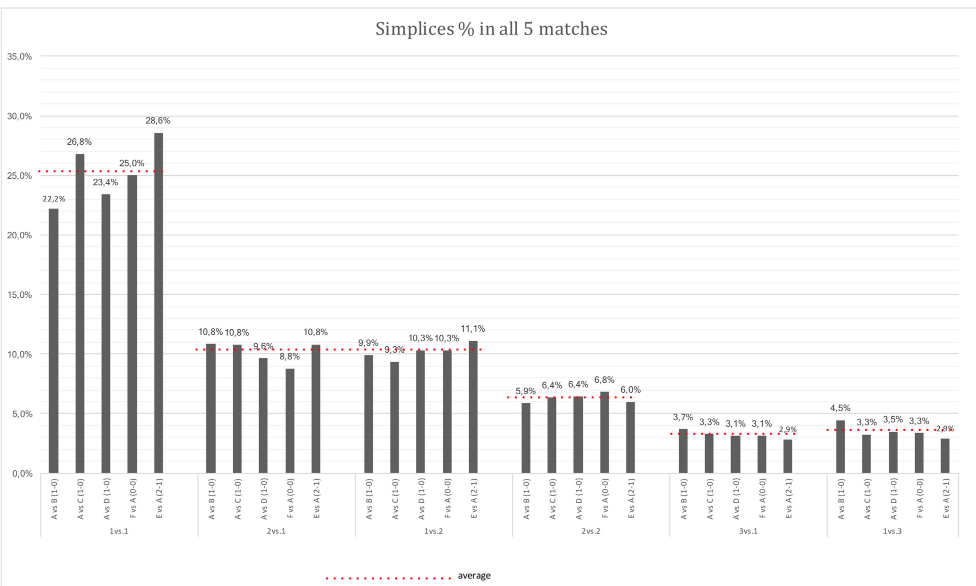


FIGURE 2 | Histogram for the most frequently occurring simplices structures in the 5 matches: 1vs.1; 2vs.1; 1vs.2; 2vs.2; 3vs.1; 1vs.3. The matches (and score) were: Team A vs. team B (1–0); Team A vs. Team C (1–0); Team A vs. Team D (1–0); Team E vs. Team A (2–1); and Team F vs. Team A (0–0).

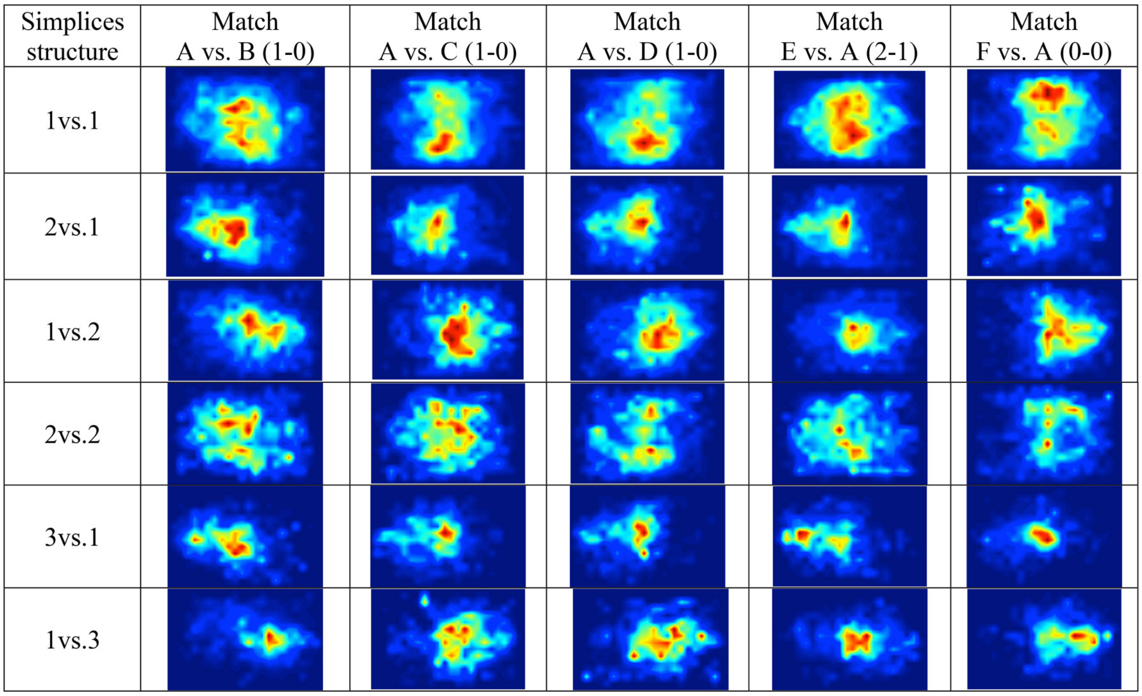


FIGURE 3 | *Heat maps* for field position of the most frequent simplices structures during the matches (when Team A, playing at home, attacks are represented from left to right). The color gradient from red to blue represents the frequency of simplices in that location (from most frequent, red, to not occurring, dark blue).

the team with the lowest score shows a tendency for a decrease in frequency of 2vs.2 near its own goal. Moreover, the next most frequently occurring simplices, 3vs.1 and 1vs.3, can be found more often close to the goal of the wining team.

Level $N + 2$ captures simplices dynamics, for example, before changes in the score. Here we present an analysis of the simplices having their geographical center closer to the goal. To answer the question "what creates an opportunity for the attackers to score?" simplices reveal how the defenders' local dominance is broken by the attackers. **Figure 5** shows an example of local dominance, in which team A (playing at home against B) scores in a counter-attack sub-phase. The play was analyzed in a set of consecutive frames (at 1 Hz) that captured the simplices nearer the goal of interest. A velocity vector computed using consecutive frames was associated to each player to show aggregation or disaggregation, as a player moved toward or away from the simplices geometric center.

The example in **Figure 5** shows that, in the frames before a goal is scored, some attacking players (e.g., 6, 7, and 10) increase their speed to place themselves in a better position either to create an invitation for a successful pass or to create a scoring opportunity. On the other hand, defensive players try to maintain or reduce interpersonal distance (e.g., 16, 19, and 22). This is aligned with other studies (Fonseca et al., 2013) where it was observed that attackers tried to increase the interpersonal distance while the defenders tried to reduce it. The consequence of these moves can be captured by simplices' configuration. This is more evident if a player stays in the same simplex or moves to another simplex. Changes in players' velocity leads to break (disaggregate) or maintain (aggregate) the simplex's integrity when they move away or toward the simplex geometric center, respectively.

Level $N + 3$ indicates how simplices interact between them, thereby creating higher-order simplices. These simplices form by aggregation of Level $N + 1$ simplices based on the proximity criterion of their geographical centers (**Figure 6**). To uncover the changes in simplex structures leading to goal scoring, higher-order simplices (**Figure 6**, purple polygons) were analyzed for the frames where significant changes occurred in the Level $N + 3$ structures (simplices of simplices).

The example of *Level* $N + 3$ analysis in **Figure 6** also reveals the connections between players before a goal was scored. The simplex formed by the goalkeeper and the goal is connected with other simplices, as the goalkeeper tries to align with the closest simplice while maintaining the link with the goal. **Figure 6** also shows how the simplices furthest from the goal are connected with simplices more directly involved in the attacking phase (i.e., closest to the goal). Other information that can be extracted from *Level* $N + 3$ is how fast changes in the link with the goal can occur, and which simplices are "disconnected," for example, on one side of the field.

DISCUSSION

The different levels of analysis of a hypernetwork can capture various degrees of team behavior dynamics, from player, to simplices, and to interactions between simplices across space and time.

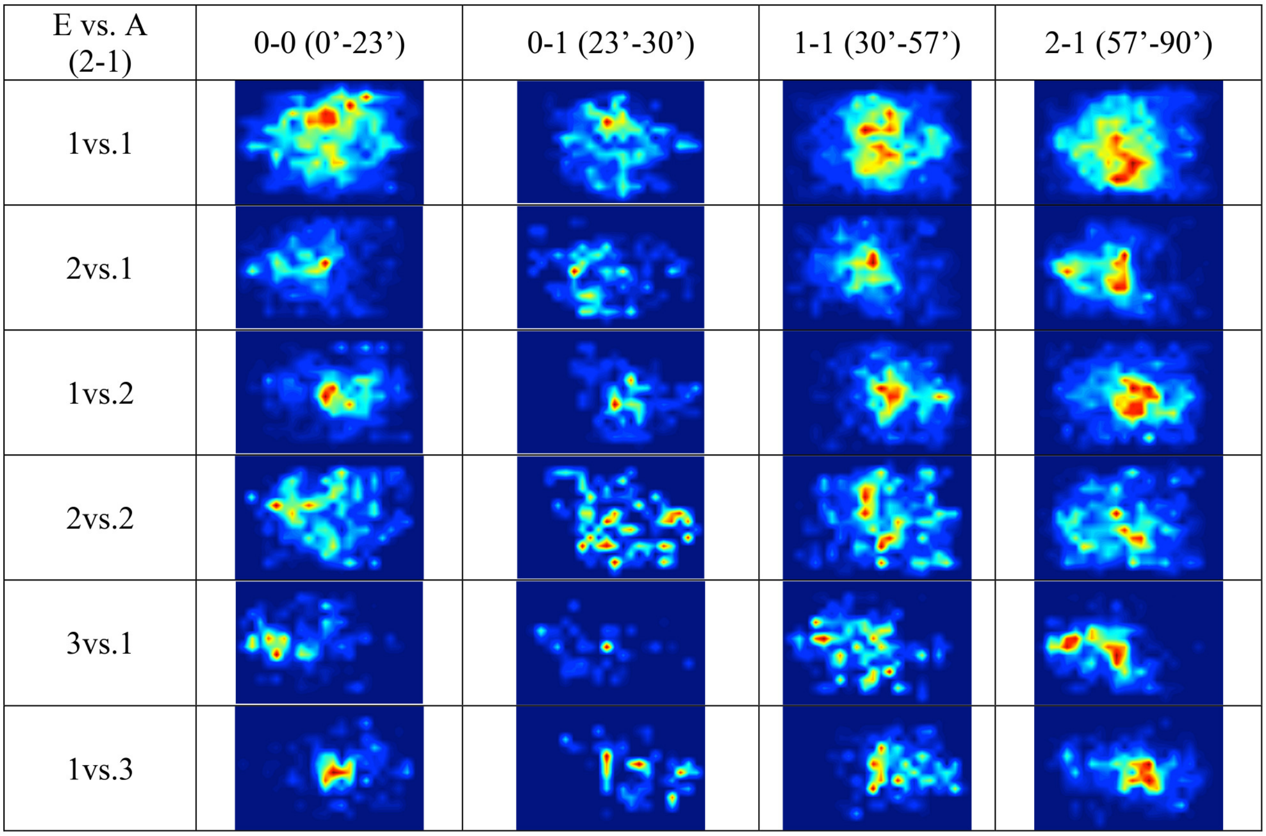


FIGURE 4 | *Heat maps* for the field position of the different simplices structures (visiting team A attacks from the right to the left hand-side). Each column corresponds to a temporal section of the match bound by a score change. The color gradient from red to blue represents the frequency of simplices in that location (from most frequent, red, to not occurring, dark blue).

At *Level N + 1*, we could identify the types of simplices occurring more often in a match, independently of their score or context (home or away). The most frequently occurring simplex was 1vs.1, followed by 1vs.0 and 0vs.1. The latter represents the link between the goalkeeper and the goal. Also occurring frequently were simplices with an unbalanced number of players, 2vs.1 and 1vs.2 (~10%), followed by the 2vs.2 simplices (~6%), and finally by the 3vs.1 and 1vs.3 simplices (~3%).

Important interpretations can be inferred from the simplices at *Level N + 1* when space and time, or contextual variables (home or away match) are considered. For example, team A won three home matches (all with score 1–0) but tied (score 0–0) or lost (score 2–1) in away games. The 1vs.1 simplices tend to occur in the mid-field and on the right of the attacking direction of team A (**Figure 3**). However, in the match lost against team E, 1vs.1 simplices were more dispersed and toward the left side of the pitch. Another frequently occurring simplex with a balanced number of players was 2vs.2, for both teams (**Figure 3**). Interestingly, these simplices also had a unique distribution in the match lost against team E, as they occurred more toward the center of the pitch and the opponent middle field. Additionally, these structures differed from match to match, showing the emergent properties of complex adaptive systems, specifically the context dependency (opponents and scoring evolution; Araújo and Davids, 2016).

Concerning simplices with an unbalanced number of players, 2vs.1 occurred more often in the center of the pitch and in the opponent middle field (similarly to 2vs.2 in the match lost against team E). The 1vs.2 simplices were also detected more often in the middle fields. Simplices 3vs.1 were distributed in the center of Team A's middle field, however, in the match against team E, they were more distant from their own goal (in the middle field). In the opposite way, in the matches against teams B and F, there were some notable occurrences of 3vs.1 simplices near team's A goal. Moreover, in these matches, 1vs.3 occurred near the center but more toward team A's middle field, suggesting that team B and F "forced" team A players away from their goal.

The results obtained considered both geographical placement and context dependency, and showed that the use of simplices formation captured match properties, such as, local dominance. These properties emerge in each match event resulting from the local interaction between players of both teams. Multilevel hypernetworks proved to be a useful method in answering to chief problems such as, the relation among micro (e.g.,

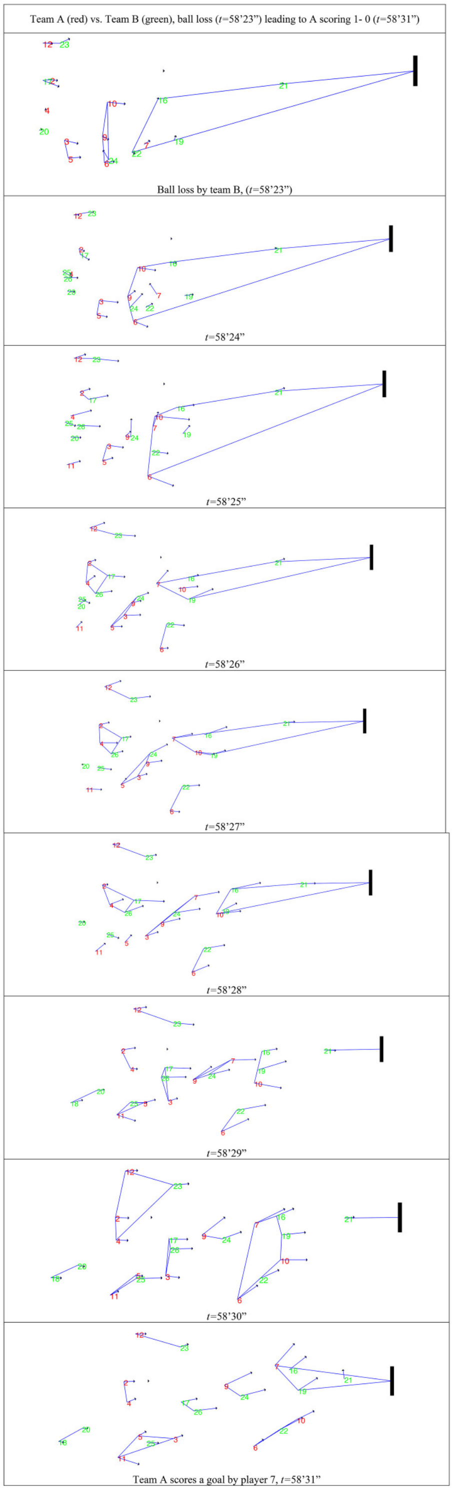


FIGURE 5 | Simplices in a sequence of nine frames (58′23″ to 58′31″) leading to a goal by Team A. Visiting players are attacking from right to left (represented in green), while home players are attacking from left to right (represented in red, including the opponents' goal). A simplex is represented by the polygon (or a line when there are only two players) defining the convex hull (or envelope) that links the nodes (players or goal). A velocity vector for each player is also presented.

players' positions), meso (e.g., local dominance), and macro levels (e.g., match result). Moreover, the use of hypernetworks allows that the analysis can consider more than the typical (in SNA) *2*-ary relations between players. These contributions fulfill previous gaps in interpersonal coordination research (Passos et al., 2016).

The analysis of the dynamics of simplices interactions at *Level* $N + 2$ revealed abrupt changes in the speed and direction of player vectors near the goal. These changes showed a tendency to be associated with transformations in simplex structure, for example, when an attacker passed through the defenders to score, or when a player disconnected from one simplex to interact with another (to balance or unbalance the simplex). The example in **Figure 5** analyzed a change in the score that resulted from a ball lost by team B in team A's middle field that led to a successful counter attack (with a goal scored). This event was characterized by transformations in the simplices' structure occurring within the short duration of the counter attack (9 s, from 58′23″ to 58′31″). Next we present the set of simplices (σ) and their evolution for these 9 s leading to a goal being scored by Team A (at 58′31″). Simplices containing the player who scored the goal are identified with (S). Simplices containing the goal are identified with (G).

$$\sigma_{1,\,58'23''}\left\langle a_3,\ a_5\right\rangle + \sigma_{2,\,58'23''}\left\langle a_9,\ a_6, a_{10}, d_{24}\right\rangle$$
$$+\,\sigma_{3,\,58'23''}\left\langle a_7,\ d_{22}, d_{16}, d_{19}, d_{21};\,(G,S)\right\rangle$$
$$\sigma_{1,\,58'24''}\left\langle a_3,\ a_5\right\rangle + \sigma_{2,\,58'24''}\left\langle a_9,\ a_6, a_{10}, d_{24}, a_7,\ d_{22}, d_{16}, d_{19}, d_{21};\,(G,S)\right\rangle$$
$$\sigma_{1,\,58'25''}\left\langle a_3,\ a_5\right\rangle + \sigma_{2,\,58'25''}\left\langle a_9,\ d_{24}\right\rangle$$
$$+\,\sigma_{3,\,58'25''}\left\langle a_6, a_{10}, a_7,\ d_{22}, d_{16}, d_{19}, d_{21};\,(G,S)\right\rangle$$
$$\sigma_{1,\,58'26''}\left\langle a_3,\ a_5, a_9,\ d_{24}\right\rangle + \sigma_{2,\,58'26''}\left\langle a_6,\ d_{22}\right\rangle$$
$$+\,\sigma_{3,\,58'26''}\left\langle a_{10}, a_7,\ d_{16}, d_{19}, d_{21};\,(G,S)\right\rangle$$
$$\sigma_{1,\,58'27''}\left\langle a_3,\ a_5, a_9,\ d_{24}\right\rangle + \sigma_{2,\,58'27''}\left\langle a_6,\ d_{22}\right\rangle$$
$$+\,\sigma_{3,\,58'27''}\left\langle a_{10}, a_7,\ d_{16}, d_{19}, d_{21};\,(G,S)\right\rangle$$
$$\sigma_{1,\,58'28''}\left\langle a_3,\ a_7, a_9,\ d_{24};\,(S)\right\rangle + \sigma_{2,\,58'28''}\left\langle a_6,\ d_{22}\right\rangle$$
$$+\,\sigma_{3,\,58'28''}\left\langle a_{10},\ d_{16}, d_{19}, d_{21};\,(G)\right\rangle$$
$$\sigma_{1,\,58'29''}\left\langle a_3,\ d_{17},\ d_{26}\right\rangle + \sigma_{2,\,58'29''}\left\langle a_9, a_7,\ d_{24};\,(S)\right\rangle$$
$$+\,\sigma_{3,\,58'29''}\left\langle a_6,\ d_{22}\right\rangle + \sigma_{2,\,58'29''}\left\langle d_{21};\,(G)\right\rangle$$
$$\sigma_{1,\,58'30''}\left\langle a_3,\ d_{17},\ d_{26}\right\rangle + \sigma_{2,\,58'30''}\left\langle a_9,\ d_{24}\right\rangle$$
$$+\,\sigma_{3,\,58'30''}\left\langle a_6, a_7, a_{10,} d_{16}, d_{19},\ d_{22};\,(S)\right\rangle + \sigma_{2,\,58'30''}\left\langle d_{21};\,(G)\right\rangle$$
$$\sigma_{1,\,58'31''}\left\langle a_9,\ d_{24},\right\rangle + \sigma_{3,\,58'31''}\left\langle a_6, a_{10,},\ d_{22}\right\rangle$$
$$+\,\sigma_{2,\,58'31''}\left\langle a_7, d_{16}, d_{19}, d_{21};\,(G,S)\right\rangle$$

The results show that certain moves performed by the player who scored the goal (player a_7) had significant impact on some simplices transformations, for example, at instants 58′27″, 58′28″, 58′29″, 58′30″, and goal scored. Player a_{10} had an important role in promoting balance in the simplex that scored the goal (with player a_7), by maintaining defender d_{19} distant from his teammate d_{16}. Moreover, player d_{19} appeared to be facing the defender's dilemma, hesitating between defending his opponent (player a_{10}) and supporting his teammate (player d_{16}). Player d_{24} was also essential in the attack play leading

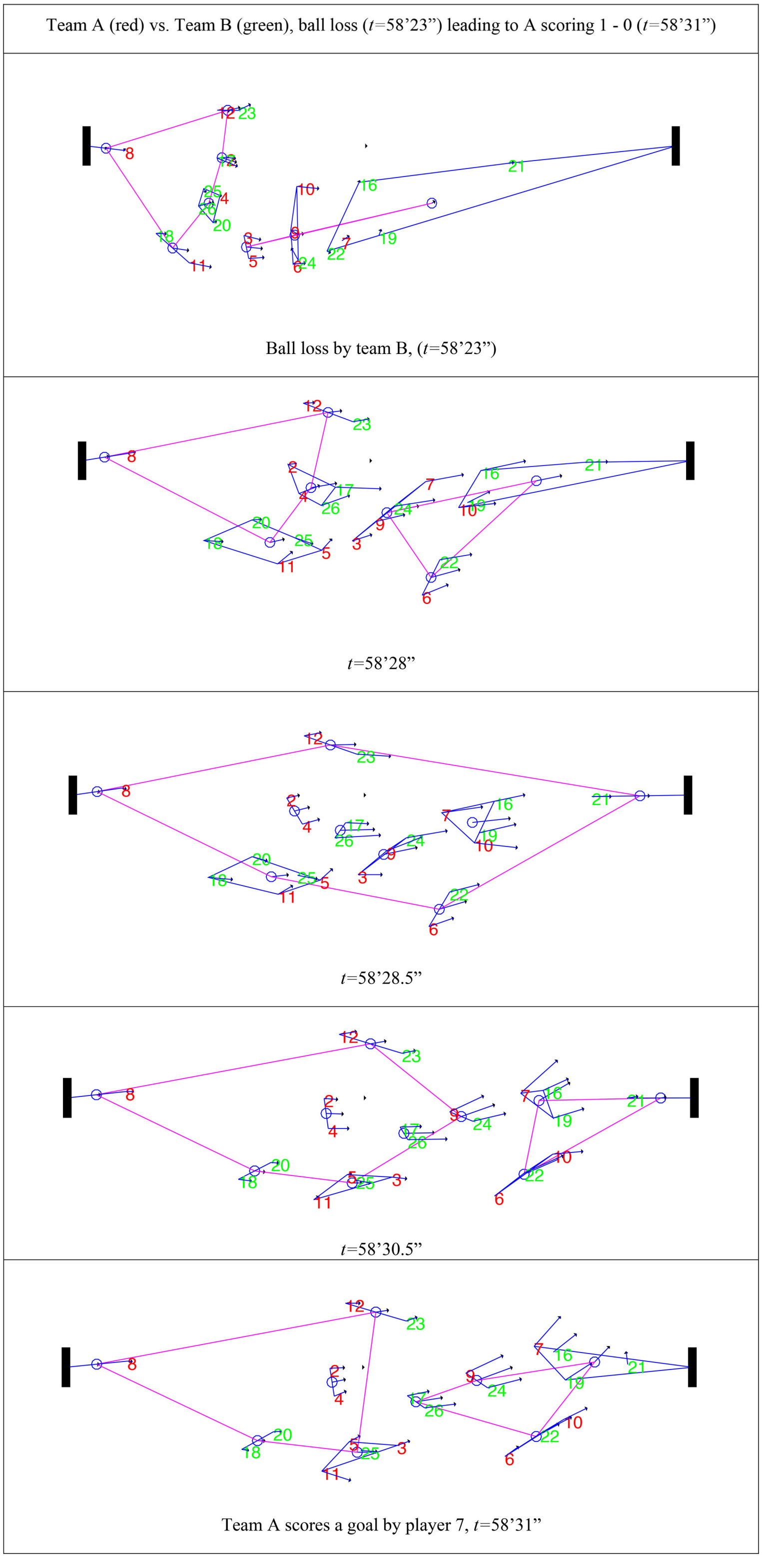


FIGURE 6 | Higher-order simplices (simplices of simplices) in a sequence of five frames before team A scores a goal. Higher-order simplices are represented by the polygon (and lines) forming the convex hull (–) that connects the geographical centers of the $N + 1$ simplices. See **Figure 5** legend for the codes for players, their velocity, and simplices.

to the goal scored, as he lost the ball but kept pursuing it, almost reaching player a_7 and thereby including him into his simplex. Finally, player a_6 broke the central simplex (containing teammate a_7) by attracting a defender toward him and hence reducing the number of players in the central middle field.

Results showed that by considering the temporal sequence of simplices transformations during critical events of the match (e.g., from ball recovery to scoring a goal) the dynamics of interaction among players is captured. Moreover, it is possible to analyze how interactions among players led to changes in simplices' structures and, consequently to such critical events (e.g., a goal scoring opportunity). Multilevel hypernetworks offer a fine temporal grain of analysis of how the micro-meso-macro level relationships emerge.

Level N + 3 clarified the dynamics of team behavior by considering the entire set of simplices, including the interactions between them (which form simplices of simplices). This level of analysis revealed the connections of players with simplices during a match. We found that the goal has an "anchoring effect" toward the goalkeeper, however, this simplex also connected with the nearer simplex (0vs.1 represents the home team and 1vs.0 the visiting team). Some simplices seemed to disconnect during critical situations, for example, when other simplices were close to the goal. This may be explained by an intentional reduction in speed by the attacking players to try and maintain the nearest defenders away from teammates (**Figure 6**).

This study showed that the hypernetworks' analysis by considering simplices of simplices reveal the degree of connection between sub-sets of players.

CONCLUSIONS AND LIMITATIONS

We have applied multilevel hypernetworks analysis, and a set of associated compound variables, to selected soccer matches by using positional variables for all players involved.

The interactions between players, as well as the sets of these interactions (simplices), were assessed based on interpersonal distance, more specifically *spatial proximity* and *instant speed* relational variables. Each player is therefore linked to his closest player (or goal, for the goalkeeper) and at higher levels, simplices are also linked to their closest simplices. The vectors representing the players' speed can represent the emergent moves from the players in order to search for new interactions or escape from others. These two "interaction variables" allowed for a deeper analysis of the structures and coordination levels emerging from the game.

Our results revealed a pattern in these interactions' dynamics that was independent of the type (home or away) and score of the match. Specifically, in every match analyzed the most frequently occurring simplices structures were, by decreasing order of frequency, 1vs.1, 2vs.1 and 1vs.2, 2vs.2, and finally, 3vs.1 and 1vs.3.

However, these simplices show differences in their distribution on the pitch, and this is particularly evident for unbalanced simplices such as, 2vs.1, 1vs.2, 3vs.1, and 1vs.3. These differential distributions are consistent with the match result (wins vs. losses) and the opponent team's strength.

We analyzed the changes in local dominance at *Level N + 2* associated with critical events (e.g., score changes) and found that dramatic speed changes can be detected in the players of simplices directly linked to the event (goal scored). Velocity is therefore the variable that allows players to improve their positioning to score or to unbalance the situation.

Finally, our last and global analysis level revealed how all the simplices were connected, but most importantly, it enabled to permanently connect all the simplices into larger hypersimplices, including the goal and goalkeeper simplex, and also the defenders and attackers who were distant from the goal.

These results may significantly contribute to improve training and playing strategies. We highlight the importance of mastering 1vs.1 situations (with and without the ball), as this structure occurs more frequently in all types of matches. For example, coaches could design exercises to train players to rapidly transform any structure into a 1vs.1 structure. Unbalanced situations such as, 2vs.1 and 3vs.1 typically reveal which team is dominating the match, particularly when those structures occur on the attacking side of that team's field. Thus, designing training exercises that create an overload for the attacking team may allow players to better adapt to such situations in a match. Finally, we found that as an attacking team moves closer to the goal, changes in player speed become more pronounced. It is therefore likely that encouraging such speed changes during training may facilitate the players' positioning inside finishing areas during a match.

Moreover, when players are connected with other players (in cooperation or competition) forming simplices, where the smaller simplices are also connected with other simplices, team coordination develops due to attunement to shared affordances and the creation of team synergies (Araújo and Davids, 2016). Training sessions may benefit from using the present analysis (e.g., most frequent cooperation/competition tag sets) and consequently design training activities that promote collective learning among groups of players (Travassos et al., 2016).

In the context of this article the criterion, closest player, for the formation of hyperedges was the only one used. The results presented at different levels of analysis are therefore conditioned and limited by this criterion. At the same time all these results where possible with only this parsimonious criterion and without any other assumptions.

Other limitation of the study is that there is no data about ball positioning, nor about "ball flux" (e.g., passes between the players). This type of interactions between players could be included by extending the proposed method with additional layers. In such layers, ball flux could be represented either as a link between players' or simplices, or alternatively as an additional term in the relationship, *R*, of the simplices.

Multilevel hypernetworks is a promising framework for soccer performance analysis that reveals important features of cooperative and competitive interactions during attacking plays. By considering space and time in multilevel analyses involving interactions between two or more players, we can obtain a richer understanding of real-world complex systems.

AUTHOR CONTRIBUTIONS

JR, main contribution regarding theoretical approach, method and results production. RL, significant contribution regarding method, software computation, and results production. PM, significant contribution on results reading and discussion and the impact to practitioners. DA, significant contribution regarding performance analysis and the general impression.

FUNDING

This work was partly supported by the Fundação para a Ciência e a Tecnologia, under Grant UID/DTP/UI447/2013 to CIPER—Centro Interdisciplinar para o Estudo da Performance Humana (unit 447).

REFERENCES

Araújo, D., and Davids, K. (2016). Team synergies in sport: theory and measures. *Front. Psychol.* 7:1449. doi: 10.3389/fpsyg.2016.01449

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., et al. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* 544, 1–122. doi: 10.1016/j.physrep.2014.07.001

Clemente, F. M., Martins, F. M. L., Kalamaras, D., Wong, D. P., and Mendes, R. S. (2015). General network analysis of national soccer teams in FIFA World Cup 2014. *Int. J. Perform. Anal. Sport* 15, 80–96. doi: 10.1080/24748668.2015.11868778

Criado, R., Romance, M., and Vela-Pérez, M. (2010). Hyperstructures, a new approach to complex systems. *Int. J. Bifurcation Chaos* 20, 877–883. doi: 10.1142/S0218127410026162

Duarte, R., Araujo, D., Correia, V., and Davids, K. (2012). Sports teams as superorganisms: implications of sociobiological models of behaviour for research and practice in team sports performance analysis. *Sports Med.* 42, 633–642. doi: 10.1007/BF03262285

Duch, J., Waitzman, J. S., and Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE* 5:e10937. doi: 10.1371/journal.pone.0010937

Fewell, J. H., Armbruster, D., Ingraham, J., Petersen, A., and Waters, J. S. (2012). Basketball teams as strategic networks. *PLoS ONE* 7:e47445. doi: 10.1371/journal.pone.0047445

Fonseca, S., Milho, J., Travassos, B., Araújo, D., and Lopes, A. (2013). Measuring spatial interaction behavior in team sports using superimposed Voronoi diagrams. *Int. J. Perform. Anal. Sport* 13, 179–189. doi: 10.1080/24748668.2013.11868640

Grund, T. U. (2012). Network structure and team performance: the case of English Premier League soccer teams. *Soc. Netw.* 34, 682–690. doi: 10.1016/j.socnet.2012.08.004

Headrick, J., Davids, K., Renshaw, I., Araújo, D., Passos, P., and Fernandes, O. (2012). Proximity-to-goal as a constraint on patterns of behaviour in attacker–defender dyads in team games. *J. Sports Sci.* 30, 247–253. doi: 10.1080/02640414.2011.640706

Johnson, J. (2006). "Hypernetworks for reconstructing the dynamics of multilevel systems," in *European Conference on Complex Systems* (Oxford, U.K).

Johnson, J. (2008). "Multidimensional events in multilevel systems," in *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach,* eds S. Albeverio, D. Andrey, P. Giordano, and A. Vancheri (Heidelberg: Physica-Verlag HD), 311–334.

Johnson, J. (2013). *Hypernetworks in the Science of Complex Systems,* Vol. 3. London: Imperial College Press.

Johnson, J. H. (2016). Hypernetworks: multidimensional relationships in multilevel systems. *Eur. Phys. J. Special Top.* 225, 1037–1052. doi: 10.1140/epjst/e2016-02653-4

Johnson, J. H., and Iravani, P. (2007). The multilevel hypernetwork dynamics of complex systems of robot soccer agents. *ACM Trans. Auton. Adaptive Syst.* 2, 1–23. doi: 10.1145/1242060.1242062

Passos, P., Araújo, D., and Volossovitch, A. (2017). *Performance Analysis in Team Sports*. London: Routledge.

Passos, P., Davids, K., and Chow, J. Y. (2016). *Interpersonal Coordination and Performance in Social Systems.* London: Routledge.

Passos, P., Milho, J., Fonseca, S., Borges, J., Araújo, D., and Davids, K. (2011). Interpersonal distance regulates functional grouping tendencies of agents in team sports. *J. Mot. Behav.* 43, 155–163. doi: 10.1080/00222895.2011.552078

Riley, M., Richardson, M., Shockley, K., and Ramenzoni, V. (2011). Interpersonal synergies. *Front. Psychol.* 2:38. doi: 10.3389/fpsyg.2011.00038

Rio, K. W., Rhea, C. K., and Warren, W. H. (2014). Follow the leader: visual control of speed in pedestrian following. *J. Vis.* 14, 1–17. doi: 10.1167/14.2.4

Schmidt, R., Bienvenu, M., Fitzpatrick, P., and Amazeen, P. (1998). A comparison of intra-and interpersonal interlimb coordination: coordination breakdowns and coupling strength. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 884–900. doi: 10.1037/0096-1523.24.3.884

Schmidt, R. C., Fitzpatrick, P., Caron, R., and Mergeche, J. (2011). Understanding social motor coordination. *Hum. Mov. Sci.* 30, 834–845. doi: 10.1016/j.humov.2010.05.014

Travassos, B., Bourbousson, J., Esteves, P. T., Marcelino, R., Pacheco, M., and Davids, K. (2016). Adaptive behaviours of attacking futsal teams to opposition defensive formations. *Hum. Mov. Sci.* 47, 98–105. doi: 10.1016/j.humov.2016.02.004

Conflict of Interest Statement: PM is affiliated with the commercial company, City Football Services. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewers AL and MD declared their shared affiliation with the handling editor, and the handling Editor states that the process nevertheless met the standards of a fair and objective review

APPENDIX

Algorithm 1 Build the simplex hyperedge set, S, from the set of all nodes, V

Require: Every node in V has one and only one node that is closest to it, $d_{Euc}(u, v)$ is the euclidean geographical distance beteeen nodes u and v

procedure BuildHyperedgeSet(V)
    $S \leftarrow \emptyset$ ▷ S, set of hyperedges build so far
    $Q \leftarrow V$ ▷ Q, set of nodes not yet in a hyperedge
    while $Q \neq \emptyset$ **do**
        $u \leftarrow i : i \in Q$ ▷ Get a focal node, u, not yet in a hyperedge
        $v \leftarrow j : d_{Euc}(u, j) \leq \min_{k \in V \setminus u} d_{Euc}(u, k)$ ▷ Get node, v, closest to u
        for each $\sigma : \sigma \in S$ **do**
            if $v \in \sigma$ **then** ▷ If closest node, v, is already in a hyperedge, σ
                $\sigma \leftarrow \sigma \cup \{u\}$ ▷ Node u is added to hyperedge σ
                $Q \leftarrow Q \setminus u$ ▷ Node u is done with
        if $u \in Q$ **then** ▷ If node u was not added to a simplex
            $\sigma_{new} \leftarrow \{u, v\}$ ▷ Create hyperedge, σ_{new}, with nodes u and v
            $S \leftarrow S \cup \{\sigma_{new}\}$
            $Q \leftarrow Q \setminus u$ ▷ Node u is done with
            $Q \leftarrow Q \setminus v$ ▷ Node v is done with
    return S

FIGURE A1 | Pseudocode for building the simplex hyperedge set.

3.3 The Interaction Between Soccer Teams Reveal Both Design and Emergence: Cooperation, Competition and Zipf-Mandelbrot Regularity

3.3.1 Context and summary

To find complexity in soccer matches through the lens of interactions (MHA) resulted in the identification of some emergent properties of complex systems, specifically from complex social systems. One of these properties, are the regularities and statistical properties found in players' sets established (spatial proximity between players) during entire matches. Complex systems empirical studies in these statistical distributions of number of items (e.g., words in texts, people in cities, tree patch sizes) have shown that these scaling properties follows empirical laws known as Zipf-Mandelbrot. Our paper demonstrated that the (re)occurrence of pitch location based sets of players in a soccer match also obeys this empirical laws. We used experimental data collected from 10 matches of the 2010/11 English Premier League that seems to be the case for most of the sets of players. The other property was revealed through the exceptions (sets that occur significantly more than the majority of the other ones) to these ZM regularities. We have found that the sets that are most frequent, corresponds to two some specific sets arrangement, expressing the artificial feature of complex social systems known as Design. The first example reveals the narrow and specific purpose of the two goalkeepers from each team, which attract them to their goals. The second example identifies in some matches the effect of symmetric spatial positioning (e.g., left defender and right attacker from opposite teams) resulting in the significantly higher occurrence of some of these sets revealing pre-defined intentionality. Therefore, the coaching process could be seen as a design discipline when implementing strategy and teams' tactics.

The Interaction Between Soccer Teams Reveal Both Design and Emergence: Cooperation, Competition and Zipf-Mandelbrot Regularity

João Ramos[1,2,*], Rui J. Lopes[1,3], and Duarte Araújo[4]

[1]ISCTE-Instituto Universitário de Lisboa, Lisbon, Portugal
[2]Universidade Europeia, Laureate International Universities, Estrada da Correia, 53, 1500-210 Lisboa, Lisbon, Portugal
[3]Instituto de Telecomunicações, Lisbon, Portugal
[4]CIPER, Faculdade de Motricidade Humana, Universidade de Lisboa, Cruz Quebrada Dafundo, Portugal
[*]jpdrs@iscte-iul.pt

ABSTRACT

Considering soccer matches as complex systems facilitates the identification of emergent properties that result from players' interactions. Such properties are the regularities and statistical characteristics found in players' couplings and sets established during the matches. Empirical studies about the statistical distributions of number of items (e.g., words in texts) have shown that these distributions follow scaling properties according to empirical laws known as Zipf-Mandelbrot. In this paper we investigate if the (re)occurrence of pitch location of sets of players in a soccer match also obeys to these empirical laws. From data collected from 10 soccer matches, results indicate that it seems to be the case for most of the sets of players. Exceptions to this are found in the sets that are most frequent and that correspond to particular types of sets (e.g., goalkeeper and goal, left defender and right attacker from opposite teams). Rather than challenging the hypothesis of a Zipf-Mandelbrot type law for this system, we argue that these exceptions are typical of design (a trait of human interaction with complex systems). Therefore, these exceptions can be explained by the elements (players) configuration design, expressing match strategy, before the team enters in such dynamical processes (the game).

Introduction

The study of complex systems has taken many approaches, a common one is to verify if a given system exhibits well known complexity features. For instance, in social complex systems there are key features such as emergent behavior resulting from self-organization. Self-organization is due in most cases to the interaction of the multiple parts of the system. One interesting and extensively investigated aspect of self-organization is the emerging exchanges of information (e.g. verbal and non-verbal communication and their statistical properties) between the people in interaction in a given system[1]. Typically these communication processes are based on synergistic relations (cooperative based interactions) and also on non-cooperative interactions such as confrontation.

In this paper we investigated how these processes and interactions are expressed in team sports. In particular, we investigated if soccer matches express similar and hallmark features present in other complex systems. Also, we addressed the influence of pre-defined design in the cooperative and competitive interactions between players. Notably in team sports, there is explicit inter-dependency between players-opponents' behavior. Often, the communication processes in soccer matches are visually-based and expressed via players' moves and interpersonal spatial relationships. In this self-organized behavior, one key feature is the synchronization of players by means of being perceptually linked due to spatial proximity. The proximity-based sets thus formed may have different dimensions, both in terms of the number of players in the set, and their constitutions, i.e., the team that each player belongs to. Each set and its inter-relationships form what is, in the context of hypernetwork theory, called a *simplex* – plural, *simplices* – of players, representing the *n-ary* spatial interactions between at least two spatially connected players[2,3]. In figure 1 it is possible to observe the simplices that are found at a particular time frame $t = 00m : 10s$ in a soccer match. In the present study each simplex is represented by the spatial convex hull enclosing the players in the set. One such simplex, simplex σ_{35}, is composed of players 3, 4 and 11 from team A (blue) and players 18 and 20 from team B (red), thus forming a 3 vs. 2 simplex[3,4]. By using the temporal aggregate of the geometrical center for each simplex convex hull a spatial histogram for that simplex can be obtained. Figures 1a and 1b illustrate two of these histograms using spatial heat maps, respectively for simplex σ_1 and σ_{35}. Simplices σ_1 and σ_5 represent a special type of relationship, that between a player, the goalkeeper, and the goal.

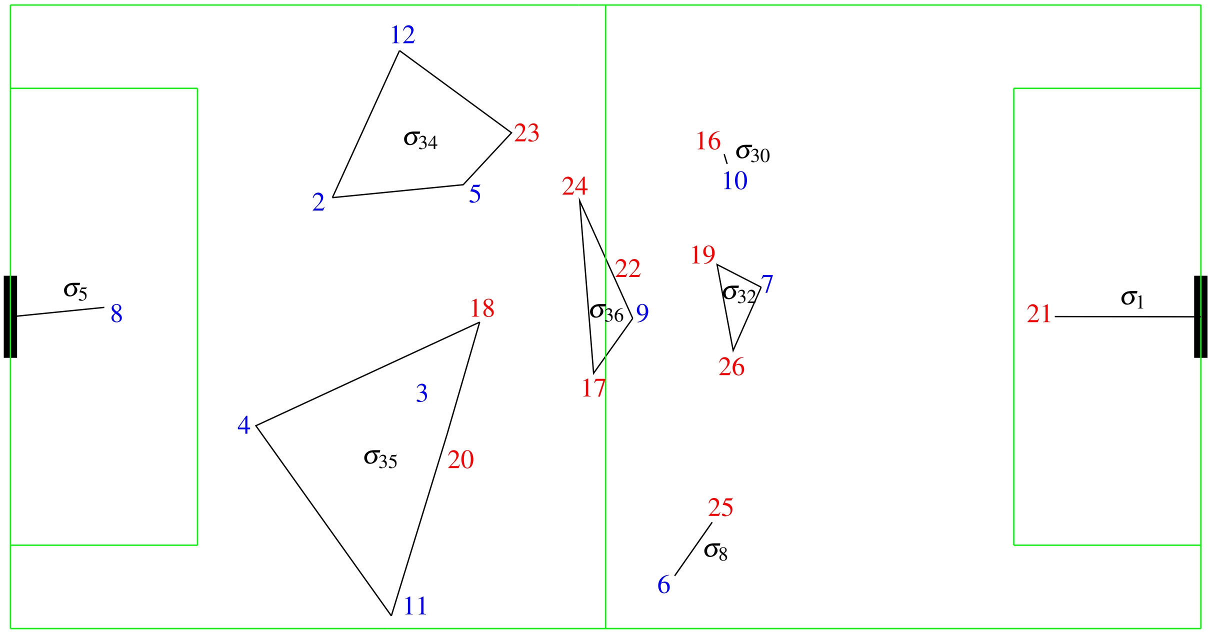
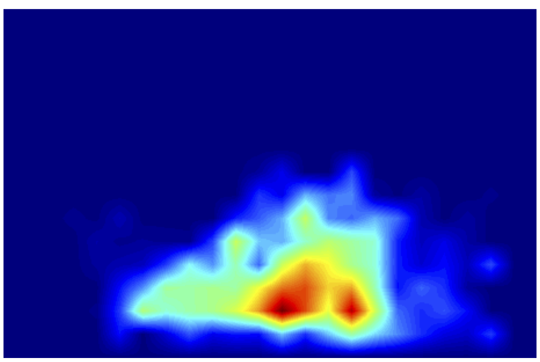
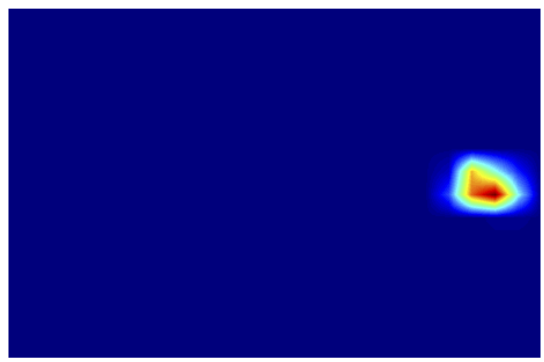


Figure 1. Players' location and simplices at frame $t = 00m : 10s$



(a) Simplex σ_{35} (3vs.2)



(b) Simplex σ_1 (B goalkeeper and goal)

This paper aims to verify if, and why, the histograms of soccer matches exhibit the scaling properties of other human and natural phenomena, usually described by power law type models. These power laws are common signatures of chaotic processes that are at one point self-organized, as it happens in many natural and social systems. Typical examples are found in the context of population distribution in big cities[5–7], forest fires[7], forest patch sizes[8], scientific citations[7,9,10], WWW surfing[7], ecology[5,7,11], solar flares[7], economic index[7], epidemics in isolated populations[7], among others. An example of these power laws is the Zipf empirical law and its generalization by Mandelbrot. In verbal communication processes such as natural language and written texts, several studies have shown that words' frequency of occurrence follows this particular type of power laws. This results from the observation that texts and languages' corpora have few words that are very frequent (e.g. "a", "the", "I", etc.) and many words that seldom occur. In Zipf's empirical law model, given the item (e.g., word) frequency, $f(r)$, and order by their assigned rank, r in decreasing order (rank 1 is for the most frequent word; rank 2 is for the second most frequent word, ...[11]), their occurrence frequency decays linearly as the rank increases on a double logarithmic scale, as expressed in equation 1

$$f(r) \propto \frac{1}{r^{\alpha}} \tag{1}$$

The generalization of this law, with a better fit to empirical data, due to Mandelbrot[12] is expressed by equation 2

$$f(r) \propto \frac{1}{(\beta + r)^{\alpha}} \tag{2}$$

Being a generalization, the later is also referred to as the Zipf-Mandelbrot (ZM) law.

In soccer performance features, such as the goal scoring distribution, also exhibit these statistical regularities related to power laws. Notably, by computing the goal's distribution in several main league soccer championships such as Brazil, England, Italy and Spain it has been found that there are very few top-scorers and many players that score only a few goals[7]. The current study focuses neither on these performance metrics nor on the individual behavior of players but rather on their relationships as expressed by the simplices' sets. That is, it investigates the systems' meso-scale properties. Questions addressed at this level typically concern processes[2,13]; thus, aiming to understand what leads to a particular simplices' set occurrence distribution[14].

In the language realm several papers addressed the question if Zipf's empirical law could not be observed in purely random systems[11,15]. These studies investigate the processes that may lead to these particular statistical distributions. In this paper we tackled a similar question but from a different angle: the co-design expressed in the match strategy. Despite the uncertainty of the human collective behavior, and thus the impossibility to predict the future state of complex systems, the deliberate design of the social structures that composes the system can promote the prevalence of some specific desired behaviors[4,16–18]. This design is in most cases a collaborative or cooperative process[19].

In soccer matches one can consider that the artificiality, i.e. the design expressed via strategic behaviour, of these social complex systems is related to the specific outcomes to be achieved[4,16,18]. In team sports a very relevant aspect of the coaching process is the implementation of the design[20]. What is particularly challenging in the study of soccer matches as social systems is that the design results from both cooperative and competitive interactions[17]. The most frequent simplices, those that seem to persist over the entire match, should thus be a consequence of this design, i.e., the teams' strategy. When each team distributes their players in the pitch (i.e., the team strategy; considering attacking, defending, midfielders, goalkeepers and left, right or in the center of the pitch), they naturally become near to symmetric players from the other team. One such example is the right attacker from team A vs. the left defender from team B. As these positions may be sustained most of the time in the match, this will lead to opposing players establishing a set of one player from team A vs. one player from team B (1 vs. 1) that occurs very frequently in the match. The formed sets may also depend on the pitch area, such as: i) the simplex set <Goalkeeper, Goal > near the goal, corresponding to players with a very specific and narrow purpose[17]; ii) the defending team trying to have numeric supremacy closer to its box (e.g., simplex σ_{32} in figure 1)[17].

The chief questions addressed in this paper, thus, refer to the simplices' set occurrence distribution, notably: i) if well known models and empirical laws, such as the Zipf-Mandelbrot law (ZM) fit the empirical distributions obtained from soccer matches; and ii) the possible impact of design, i.e. match strategy, on the simplices' set statistical distribution.

Methods

Raw data: players' coordinates

The raw data used in this study consisted in the 22 players' two-dimensional displacement coordinates provided by PROZONE (now STATS) for 10 matches, five at home and five away, of a focus team (team A) in the 2010/2011 English Premier League Season. This data was obtained via a semi-automatic tracking system based on multiple-camera analysis. In the system provided by PROZONE the position of the 22 players during the match is estimated based on the synchronized video files from eight cameras placed on the top of the stadium operating at a frequency of 10Hz (i.e., 10 frames per second, producing about 54000 frames per match)[21]. Player's substitutions and sent-offs are also considered using ancillary descriptions of the match, e.g., commentary metadata.

Building of simplices' sets and their heat maps

For each frame in the match the typical 22 players in the pitch plus the two goals are organized in sets, the simplices sets, according to the computational procedure adopted by Ramos and colleagues[3]. The criteria for selecting the players (or goals) for each set are based only in spatial proximity. In this paper the two goals are also considered in the simplices formation as they act as special spatial references to the players, namely the goalkeepers. Figure 1 illustrates the players' and goals' position in a particular time frame, were players and goals in the same simplex set are connected within their convex hull. Each simplex is uniquely defined by its index, i and by its element set, σ_i, such that: $\sigma_i = \sigma_j \implies i = j$. For each frame, t, Σ_t is the set of all simplices' sets that are found in that frame.

ZM model, ranking and bootstrapping the simplices sets

Zipf and Mandelbrot empirical laws relate token values and their rank using, respectively, equation 1 and 2. These laws can also be expressed by a probability density function 3.

$$p_{\theta,r} = \frac{C}{(\beta + r)^{\alpha}} \tag{3}$$

Where $p_{\theta,r}$ is the probability density value for token with rank r under parameter set $\theta = \{\alpha, \beta\}$. The value of C is given by equation 4.

$$\sum_{r=cut}^{s} p_{\theta,r} = 1 \tag{4}$$

The upper limit of the summation in 4 is s, which is the number of different simplices observed in the entire match. On the other hand, given that in this study we also investigate the impact of design in the most frequent simplices sets we use also left truncation in the generalization of these probability density function. Correspondingly, the summation lower limit, cut, defines the rank used to left truncate the distribution. For example, if $cut = 1$ all simplices sets are considered, corresponding to the usual case in the literature, if $cut = 3$ the two most frequent simplices sets are not considered.

Using a counting process, computed over the entire match, we obtain the frequencies for each simplex set, n_r. These frequencies are used to rank the simplices. This counting process is defined in equations 5 and 6

$$n_r = \sum_{t=1}^{T} I_r(t) \tag{5}$$

where T is the number of samples in the match and $I_r(t)$ is an indicator function given by:

$$I_r(t) = \begin{cases} 1, & \sigma_r \in \Sigma_t \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The total number of simplices' counts, n is given by:

$$n = \sum_{k=cut}^{s} n_r \tag{7}$$

where n_r is obtained from equation 5 and the summation upper and lower limits are the same as in equation 4. In order to avoid the artifacts resulting from using the same data set for both ranking and frequency value described by Piantadosi[12] we used a bootstrapping procedure similar to the one proposed also by Piantadosi[12]. The bootstrapping process is also used for defining confidence intervals for the frequency values[22] and for assessing the ZM law fit to the empirical data.

Fitting and validating the ZM distribution model

The analysis of these data structures related to the ZM distribution in real-life situations, implies an effective fitting procedure and an appropriate test for the goodness of fit[5]. The estimation for the unknown parameters can be obtained by applying a Maximum Likelihood Estimation (MLE).

The likelihood estimator l_θ for ZM is given by equation 8:

$$l_\theta = \frac{n!}{n_{cut}! \cdot n_{cut+1}! \ldots n_s!} \cdot \prod_{r=cut}^{s} p_{\theta,r}^{n_r} \tag{8}$$

Taking the logarithm of l_θ:

$$L_\theta = \ln\left(\frac{n!}{n_{cut}! \cdot n_{cut+1}! \ldots n_s!}\right) + n\ln(C) - \alpha \sum_{r=cut}^{s} (n_r(r+\beta)) \tag{9}$$

where s is the number of different simplices observed, and n_r is the observed number of occurrences for simplex with rank r.

The estimation of the values for parameters α and β that minimize $-L_\theta$ is performed using the numerical minimization provided by Octave's package *optim* function *fminsearch* via the Nelder & Mead Simplex algorithm[23]. Parameter C is obtained from equation 4.

To test the validity of the model we used the χ^2 metric for assessing its goodness of fit[24–26]. Although the $p-value$ obtained from the χ^2 statistic is used to decide if the hypothesis should or not be rejected, we decided to show the χ^2/n value as it does not depend on the sample size and where the "rule of thumb" $\chi^2/n < 1$ can be applied.

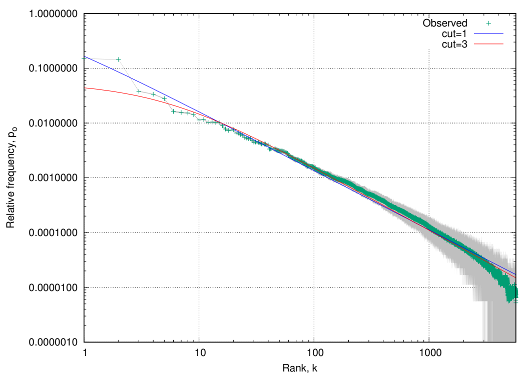
Results

We present the results of the simplices formation over the entire 10 soccer matches; in particular, the following figures show the specific results from three selected matches opposing: team A against team B (figure 2), team A against team C (figure 3) and team A against team H (figure 4). In the figures: 2a, 3a and 4a, where we plotted the simplices' set relative frequencies versus the rank from the observed data. The gray area in these figures is obtained via a bootstrapping process where the limits correspond to the 10% and 90% percentiles (as described in subsection Bootstrapping). The red and blue lines in sub-figures 2a, 3a and 4a correspond to the values obtained from the ZM model, for $cut = 1$ and $cut = 3$, with parameters C, α and β estimated via Maximum Likelihood Estimation (MLE)[5].

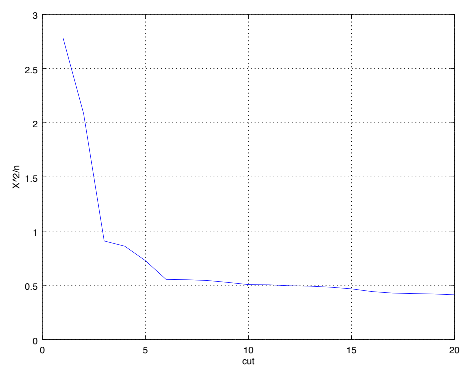
In sub-figures: 2b, 3b and 4b we plotted the χ^2/n metric for assessing the goodness of fit[24–26] of the ZM model. (We opted for plotting the χ^2/n instead of χ^2 as it is easier to identify the $\chi^2/n < 1$ rule of thumb criteria for not rejecting the hypothesis.) This metric is computed according to expressions 8 and 9 and plotted against the *cut* value.

Sub-figures c), d) and e) for all three cases show the simplices' position heat maps for the 3^{rd} to 5^{th} most frequent occurring simplices (i.e., we do not show the <Goalkeeper, Goal > simplices sets, as further explained in the Discussion), for the entire corresponding matches. Finally in table 1 we present the parameter values (β, α and χ^2/n) of the Mandelbrot generalization for all the 10 matches considered. These results are obtained considering two different conditions: considering all the existing simplices (*cut*=1); and removing the two most occurring simplices (*cut*=3).

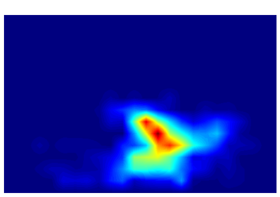
Figure 2. Team B vs. Team A



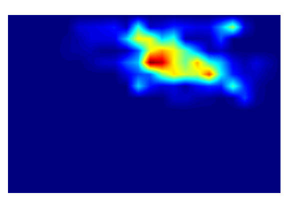
(a) Simplices sets relative frequency



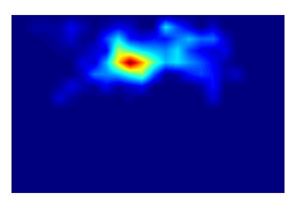
(b) Model goodness of fit (χ^2/n)



(c) Simplex 25 (1vs.1)



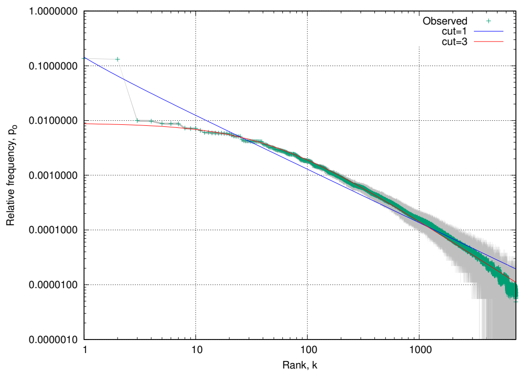
(d) Simplex 49 (1vs.1)



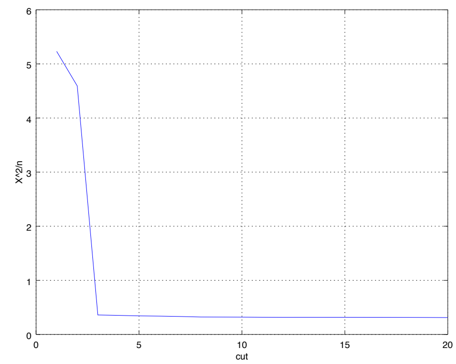
(e) Simplex 54 (1vs.1)

The results shown for the observed data in figure 2a suggested that the frequency versus rank follows a power law (we used as hypothesis the ZM model). The results shown in figure 2b, where χ^2 is used to assess the goodness of fit of the ZM model, lead to different conclusions depending on how many simplices sets are considered. In the case where all simplices sets are considered ($cut = 1$, depicted in figure 2a as a blue line) the ZM model hypothesis must be rejected. The high value for the χ^2 statistic is mostly due to the most frequent simplices sets that clearly do not follow a power law as they form groups with very similar (and high) frequencies. The results in figure 2b show that the χ^2 statistic decreases with the *cut* value and that for $cut \geq 3$ the ZM hypothesis should not be rejected. In figure 2a we represent in the red line the ZM model which results for this threshold ($cut = 3$). A notable difference between when using the ZM model for the two thresholds ($cut = 1$) and ($cut = 3$) is found in the β parameter, that shifts from almost Zipf-like ($\beta = 0.16978$) to clearly Mandelbrot ($\beta = 4.6029$). Figures 2c to 2e show the spatial position heat maps for the simplices ranked 3^{rd} to 5^{th} in Match B vs. A. They all correspond to simplices sets formed by one player of team A and one player from team B (i.e., 1vs.1) and that are located in very particular zones of the pitch (along the side lines).

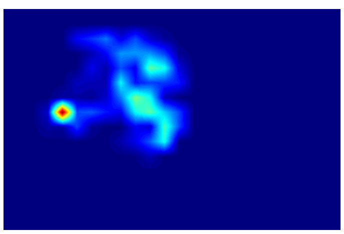
Figure 3. Team C vs. Team A



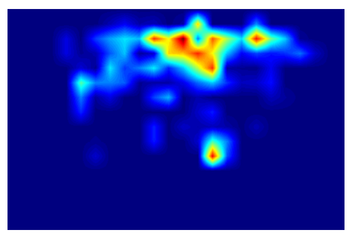
(a) Simplices sets relative frequency



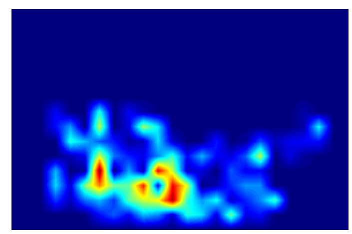
(b) Model goodness of fit (χ^2/n)



(c) Simplex 38 (2vs.1)



(d) Simplex 352 (1vs.1)



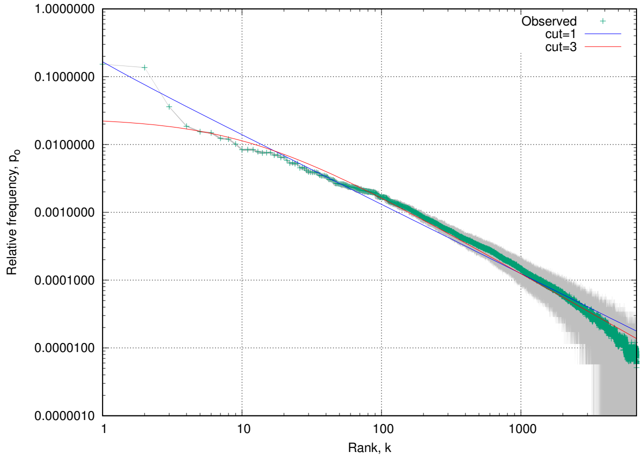
(e) Simplex 1371 (1vs.1)

Figures 3a to 3e present the same results for Match Team C vs. Team A. In the results shown in figure 3b we observe the same threshold value, $cut = 3$, for not rejecting the ZM model hypothesis. In the case of this match, this cut value is much more significant as there is no substantial change in the χ^2 value after this threshold. This is also clear in 3a where the red line exhibits a much better fit to the observed data after the 2^{nd} most frequent simplex set. Again, a notable difference is found in the β parameter when using the ZM model for the two thresholds ($cut = 1$) and ($cut = 3$), as it shifts from almost Zipf-like ($\beta = -0.20276$) to clearly Mandelbrot ($\beta = 43.603$). Figures 3c to 3e show the spatial position heat maps for the simplices ranked 3^{rd} to 5^{th} in Match C vs. A. Two of these heat maps (3d and 3e) correspond to 1vs.1 simplices along the side lines. On the other hand, the 3^{rd} most frequent simplex set (3e) corresponds to an unbalanced set (two players from Team C and one player from Team A) and the spatial position of the heat map is more intense in the central zone of the pitch and close to Team's C goal.

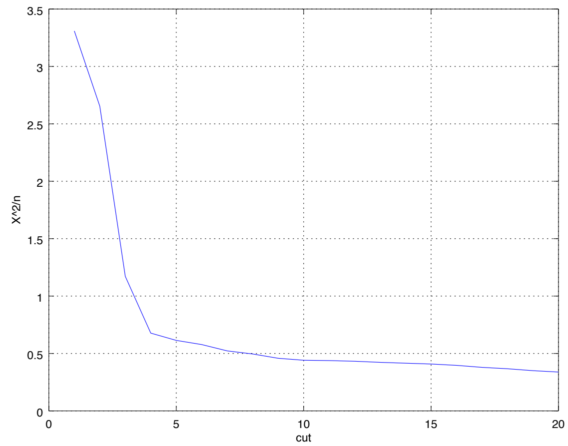
Figures 4a to 4e present the same results for Match Team A vs. Team H. In the results shown in figure 4b we observe now that the threshold value for not rejecting the ZM model hypothesis is $cut = 4$. This is also clear in 3a where the red line exhibits a much better fit to the observed data after the 3^{rd} most frequent simplex set. The notable difference for the two previous matches is that the 3^{rd} most frequent simplex set also stands out from all the others. Again, a notable difference in β parameter is found when using the ZM model for the two thresholds ($cut = 1$) and ($cut = 3$), that also shifts from almost Zipf-like ($\beta = -0.12456$) to clearly Mandelbrot ($\beta = 10.466$). Figures 4c to 4e show the spatial position heat maps for the simplices ranked 3^{rd} to 5^{th} in Match A vs. H. Two of these heat maps (4c and 4d) correspond to 1vs.1 simplices along the side lines. Figure 4c for the 3^{rd} most frequent simplex set reinforces the relevance already mentioned for this simplex set. Figure 4e corresponds to an unbalanced set (one player from Team A and two players from Team H) and the spatial position of the heat map is more intense in the central zone of the pitch and close to Team's H goal.

In the cut=1 table, the results for the β values are closer to 0, which approximates to a Zipf's like distribution. The

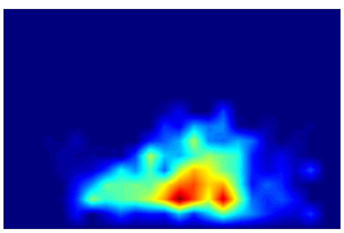
Figure 4. Team A vs. Team H



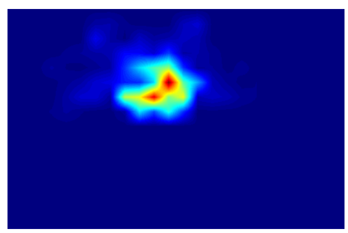
(a) Simplices sets relative frequency



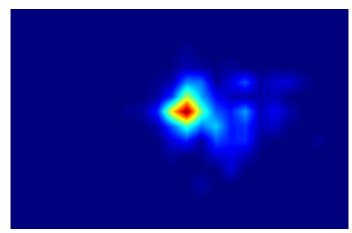
(b) Model goodness of fit (χ^2/n)



(c) Simplex 8 (1vs.1)



(d) Simplex 28 (1vs.1)



(e) Simplex 32 (1vs.2)

Table 1. Mandelbrot parameter of β and α

(a) $cut = 1$

Match	β	α	χ^2/n
B x A	0.16978	1.0808	2.7847
C x A	-0.20276	0.9763	5.2315
D x A	0.09027	1.0601	3.3324
E x A	0.09536	1.0442	3.4538
F x A	0.49587	1.1060	2.3015
A x G	0.00658	1.0536	2.6180
A x H	-0.12456	1.0217	3.3092
A x I	-0.02822	1.0419	4.0488
A x J	0.10487	1.0427	2.0488
A x K	0.07262	1.0579	3.2862

(b) $cut = 3$

Match	β	α	χ^2/n
B x A	4.6029	1.1532	0.9089
C x A	43.603	1.3058	0.3598
D x A	8.7968	1.1876	0.6676
E x A	12.5346	1.2009	0.5386
F x A	7.7196	1.2231	0.4570
A x G	5.5558	1.1347	0.9276
A x H	10.466	1.1576	1.1718
A x I	12.634	1.2019	0.7568
A x J	6.7537	1.1448	0.8013
A x K	8.7551	1.1851	1.0925

distribution begins to approximate a Mandelbrot distribution when we remove the two first most occurring simplices, where β values are significantly higher.

In the three selected matches it is possible to identify that the removal of the first two most frequent simplices improves the goodness of fit, as shown in the a) and b) figures for all cases. We can also observe that the first two simplices stand out from all the other simplices sets, not only on the fact that they have the highest frequency values, but also that when removed from the data set this results in significantly smaller χ^2 values on the goodness of fit tests[24,25]. It is also interesting to note that when these simplices are considered, the distribution is approximately Zipf (i.e., $\beta \approx 0$), whilst if not considered then the Mandelbrot generalization must be considered (i.e., $\beta > 4.5$).

These results reveal that, in the ten soccer matches analyzed, the frequency of the overwhelming majority of the simplices that emerge obeys a complex systems' typical distribution. This is supported by the goodness of fit tests that allows us not to reject the null hypothesis that the simplices frequencies follow a Zipf-Mandelbrot (ZM) like distribution[25].

Discussion

The results obtained in this study provided valuable insights about two questions. On the one hand, they revealed that for most of the simplices observed in the ten soccer matches they present a statistical distribution of occurrence that is typical to complex systems. This is supported by the goodness of fit tests on the hypotheses of Zipf-Mandelbrot (ZM) like distribution[25]. Distributions that correspond to hallmarks of complex and self-organized criticality[27]. We could also observe that the first two simplices stand out from all the other simplices sets, both on their frequency values, and on their impact on the ZM distribution *beta* parameter (from $\beta \approx 0$, Zipf, to $\beta > 4.5$, Mandelbrot).

This leads us to the second question as the players involved in the two most frequent simplices, the goalkeepers, have a very distinctive purpose (defending the goal), with specific rules, compared with the other players: i) first, these simplices sets are of the type <Goalkeeper, Goal> and the design of the competition field is established with specific delimited areas in the pitch, maybe because of the rule that allows goalkeepers to touch the ball with the hands in a specific area of the pitch; ii) second, this specific role of these players produces a kind of anchoring effect of the goalkeepers to their goals, as if they are attracted to the goal, to limit the opposing team players to connect with the goal. In this context we can observe another typical feature when dealing with social complexity, that is, there is intentionality in the behavior of the actors in the complex social systems[4, 16–18].

Notably, some of the matches (e.g. A against teams B and H), reveal other simplices that seem to be designed, preplanned or conceived before the match, to behave differently from the others, i.e., where subsets of players are more frequently close to each other than the others (figures 2a and 4a). This clearly shows in match B vs. A where there are mainly 1vs.1 simplices and their typical positioning in the field (figures c), d) and e)), during the entire match, shows that they stayed connected in a very specific area of the field. The same happens in the match A vs. H, where this also occurs for the six first more frequent simplices.

On the other hand, when we analyzed the match C vs. A we found that, with the exception of <Goalkeeper, Goal> simplices, there are no other simplices that stand out from the rank distribution. This is also expressed in figure b) where after cutting the first two already analyzed <Goalkeeper, Goal> simplices, the χ^2 values maintain lower (less than 0.5) and stable after this cut.

In conclusion, we have found many 1vs.1 simplices and their closer combinations 1vs.2 or 2vs.1 and also 2vs.2, that might reflect a preformed design and strategy of the teams, but also a more rare (less frequent) large number of sets of simplices that emerge and that reflects many interactions that are self-organized. It is interesting to note that on the one hand the frequency distribution of simplices sets is well modeled by the ZM model, a hallmark of complex systems, with α parameter in the range of other systems (e.g., written text, population size); and on the other hand the largest deviations from this model occurs for the most common simplices sets, revealing design - a well identified means to deal with complexity. This latter aspect is particularly relevant as it results not only from the traditional cooperative design[19] but in this case from both cooperative and competitive processes.

References

1. Riley, M. A., Richardson, M., Shockley, K. & Ramenzoni, V. C. Interpersonal synergies. *Front. psychology* **2**, 38 (2011).

2. Johnson, J. Multilevel multidimensional networks for designing and managing complex human systems. In *Conference on Complexity Science and Society*, 11–14.

3. Ramos, J., Lopes, R. J., Marques, P. & Araújo, D. Hypernetworks reveal compound variables that capture cooperative and competitive interactions in a soccer match. *Front. Psychol.* **8**, 1379 (2017).

4. Jeffrey, J. *Hypernetworks In The Science Of Complex Systems*, vol. 3 (World Scientific, 2013).

5. Izsák, F. Maximum likelihood estimation for constrained parameters of multinomial distributions—application to Zipf–Mandelbrot models. *CMStatistics* **51**, 1575–1583 (2006).

6. Li, G. & Wang, X. Prediction accuracy measures for a nonlinear model and for right-censored time-to-event data. *arXiv preprint arXiv:1611.03063* (2016).

7. Malacarne, L. & Mendes, R. Regularities in football goal distributions. *Phys. A: Stat. Mech. Its Appl.* **286**, 391–395 (2000).

8. Saravia, L. A., Doyle, S. R. & Bond-Lamberty, B. Power laws and critical fragmentation in global forests. *Sci. Reports* **8**, 17766 (2018).

9. Silagadze, Z. Citations and the Zipf-Mandelbrot's law. *Complex Syst.* **11**, 487–499 (1997).

10. Komulainen, T. Self-similarity and power laws. *Complex systems-science on edge chaos (Report 145* (2004).

11. Ferrer-i Cancho, R. & Elvevåg, B. Random texts do not exhibit the real Zipf's law-like rank distribution. *PLoS One* **5**, e9411 (2010).

12. Piantadosi, S. T. Zipf's word frequency law in natural language: A critical review and future directions. *Psychon Bull Rev* **21**, 1112–1130 (2014).

13. Ramos, J., Lopes, R. J. & Araújo, D. What's next in complex networks? capturing the concept of attacking play in invasive team sports. *Sports Medicine* **48**, 17–28 (2017).

14. Ramos, J., Lopes, R. J., Marques, P. & Araújo, D. Hypernetworks: capturing the multilayers of cooperative and competitive interactions in soccer. In *International Congress Complex Systems in Sport*, 150–153 (Frontiers, 2017).

15. Wentian, L. Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Inf. Theory* **38**, 1842–1845 (1992).

16. Johnson, J. *Embracing Complexity in Design*, chap. Embracing design in complexity, 193–203 (Routledge, 2010).

17. Blecic, I. & Cecchini, A. Design beyond complexity: Possible futures—prediction or design?(and techniques and tools to make it possible). *Futur.* **40**, 537–551 (2008).

18. Johnson, J. Science and policy in designing complex futures. *Futur.* **40**, 520 – 536 (2008).

19. Johnson, J. Complexity science in collaborative design. *CoDesign* **1**, 223–242 (2005).

20. Rothwell, M., Davids, K. & Stone, J. Harnessing socio-cultural constraints on athlete development to create a form of life. *J. Expert.* (2017).

21. STATS. Football performance analysis solutions (2018). URL https://www.stats.com/football/. Last accessed 26 October 2018.

22. Babu, G. & Bose, A. Bootstrap confidence intervals. *Stat. & Probab. Lett.* **7**, 151–160 (1988).

23. Nelder, J. A. & Mead, R. A simplex method for function minimization. *The computer journal* **7**, 308–313 (1965).

24. Baker, S. & Cousins, R. D. Clarification of the use of chi-square and likelihood functions in fits to histograms. *Nucl. Instruments Methods Phys. Res.* **221**, 437–442 (1984).

25. Bentler, P. & Bonett, D. G. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* **88**, 588 (1980).

26. Spiess, A.-N. & Neumeyer, N. An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC pharmacology* **10** (2010).

27. Schmidt, R. C. & Fitzpatrick, P. The origin of the ideas of interpersonal synchrony and synergies. *Interpers. coordination performance social systems. London: Routledge* (2016).

Acknowledgements

Duarte Araújo was partly supported by the Fundação para a Ciência e Tecnologia, under Grant UID/DTP/UI447/2019 to CIPER—Centro Interdisciplinar para o Estudo da Performance Humana (unit 447). Rui Jorge Lopes was partly supported by the Fundação para a Ciência e Tecnologia, under Grant UID/EEA/50008/2019 to Instituto de Telecomunicações.

Author contributions statement

J.R. conceived the original ideas presented in this article. J.R. and R.J.L. completed the model establishment and computation. J.R, R.J.L. and D.A. analyzed and commented the results. Text and figures where mostly written by J.R. All authors reviewed the manuscript.

Additional information

Competing financial interests

The authors declare no competing interests.

3.4 Hypernetworks: Capturing the Multilayers of Cooperative and Competitive Interactions in Soccer

3.4.1 Context and summary

In this study we have extended the interaction-based Multilevel Hypernetworks Approach (MHA) used in the previous presented paper in section 3.2 (Ramos, Lopes, et al., 2017a). The different, micro-meso-macro levels of analysis, allowed in each level to identify Backcloth (more stable structures) and Traffic (dynamics of those structures). Regarding Backcloth, we have proposed more complete formalisms to represent each simplex (set of players'), introducing the information about some aspects of the relationship between the players' (local dominance). This solution tackle the "intermediate word problem" (Johnson, 2006) using soccer technical terminology and becoming more intuitive for practitioners understanding. Additionally, the statistical information not only on how much simplices of each type exists at every single moment or in the entire match, but also on what specific simplex occurred more in the entire match and where did it emerged in the pitch (histograms representing heat maps). This information, is highly relevant for the understanding of which players' were more dynamic in promoting symmetry or breaking it and if the dynamics of the simplices reflects or not the strategic thinking of what was prepared for the match. Also, we can identify in each critical event, those players' moves regarding aggregation or disaggregation and how did he do it (through maintaining velocity, acceleration or direction changes).

3.4.2 Paper *authors'* copy

Hypernetworks: Capturing the Multilayers of Cooperative and Competitive Interactions in Soccer

João Ramos[1,2*], Rui J. Lopes[1,4], Pedro Marques[5], Duarte Araújo[3]

[1] ISCTE-Instituto Universitário de Lisboa, Lisbon, Portugal

[2] Universidade Europeia, Laureate International Universities, Estrada da Correia, 53, 1500-210 Lisboa, Portugal

[3] CIPER, Faculdade de Motricidade Humana, Universidade de Lisboa, Cruz Quebrada Dafundo, Portugal

[4] Instituto de Telecomunicações, Lisbon, Portugal

[5] Football Performance, City Football Services.

Introduction

Hypernetwork theory brings together the micro-meso-macro levels of analysis of interaction-based complex systems (Boccaletti et al., 2014; Johnson, 2013). This study considers team synergies (Araújo & Davids, 2016), where teams and athletes are co-evolving subsystems that self-organize into new structures and behaviors. The emergent couplings of players' movements have been studied, considering mostly the distance between a player and the immediate opponent (e.g., Headrick et al., 2012), and other interpersonal distance measures (Fonseca et al., 2013; Passos et al., 2011).

Such emergent interpersonal behavior of soccer teams can be captured by multilevel hypernetworks approach that considers and represents simultaneously the minimal structure unit of a match (called simplex). More stable structures are called backcloth. The backcloth structure that represents soccer matches is not limited to the binary relations (*2*-ary) studied successfully by social networks analysis (SNA) but can consider also n-ary relations with n>2.

These simplices are most of the times composed of players from both teams (e.g. 1vs.1, 2vs.1, 1vs.2, 2vs.2) and the goals. In a higher level of representation, it is also possible to represent the events associated, like the interactions between players and sets of players that could cause changes in the backcloth structure (aggregations and disaggregation of simplices).

The main goal of this study was to capture the dynamics of the interactions between team players at different scales of analysis (micro – meso - macro), either from the same team (cooperative) or from opponent team players (competitive).

Methods

To analyze the interactions of players, we used proximity criteria (closest player) for defining the set of players in each simplex. The non-parametric feature of this method allows for the analysis of the sets (simplices) that emerge from spatiotemporal data of players and form simplices of different types.

In this study, we first used the mathematical formalisms of hypernetworks to represent a multilevel team behavior dynamics, including micro (interactions between players established through interpersonal closest distance), meso (dynamics of a given critical event, e.g., goal scoring opportunity) and macro levels (dynamics of emerging local dominance). We have applied hypernetworks analysis to soccer matches from the English premier league (season 2010-2011) by using two-dimensional player displacement coordinates obtained with a multiple-camera match analysis system provided by STATS (formerly Prozone).

Results

We studied different levels of analysis. At the micro level, we found:

i. The most common minimal simplices are 1 vs. 1 (25.0%), followed by 1 vs. 2 (10.31%), 2 vs. 1 (8.78%) and 2 vs. 2 (6.81%);
ii. Which players were more often connected forming the same simplices (see Table 1).
iii. Where did it take place (*heat maps*) in field game (Figure 1)?

$\sigma_{49} = \langle a_5, b_{25}; (1\ vs.\ 1) \rangle$ 0,302	$\sigma_{48} = \langle a_4, b_{19}; (1\ vs.\ 1) \rangle$ 0,127	$\sigma_{63} = \langle a_4, b_{16}, b_{19}; (1\ vs.\ 2) \rangle$ 0,107	$\sigma_{171} = \langle a_2, b_{20}; (1\ vs.\ 1) \rangle$ 0,076	$\sigma_{177} = \langle a_5, b_{17}, b_{25}; (1\ vs.\ 2) \rangle$ 0,061
$\sigma_{25} = \langle a_{10}, b_{18}; (1\ vs.\ 1) \rangle$ 0,293	$\sigma_{24} = \langle a_7, b_{21}; (1\ vs.\ 1) \rangle$ 0,124	$\sigma_{96} = \langle a_3, a_{12}, b_{22}; (2\ vs.\ 1) \rangle$ 0,096	$\sigma_{331} = \langle a_2, b_{16}, b_{20}; (1\ vs.\ 2) \rangle$ 0,067	$\sigma_{240} = \langle a_6, b_{24}; (1\ vs.\ 1) \rangle$ 0,058
$\sigma_{54} = \langle a_8, b_{26}; (1\ vs.\ 1) \rangle$ 0,266	$\sigma_{182} = \langle a_3, b_{22}; (1\ vs.\ 1) \rangle$0, 121	$\sigma_{178} = \langle a_{10}, b_{18}, b_{21}; (1\ vs.\ 2) \rangle$ 0,089	$\sigma_6 = \langle a_7, a_{10}, b_{18}, b_{21}; (2\ vs.\ 2) \rangle$ 0,064	$\sigma_{408} = \langle a_{11}, b_{17}; (1\ vs.\ 1) \rangle$ 0,057

Table 1. Relative frequency for the top 15 simplices in the analyzed match. (e.g. simplex $\sigma_{49} = \langle a_5, b_{25}; (1\ vs. 1)\rangle$ was found 30.2% of the time).

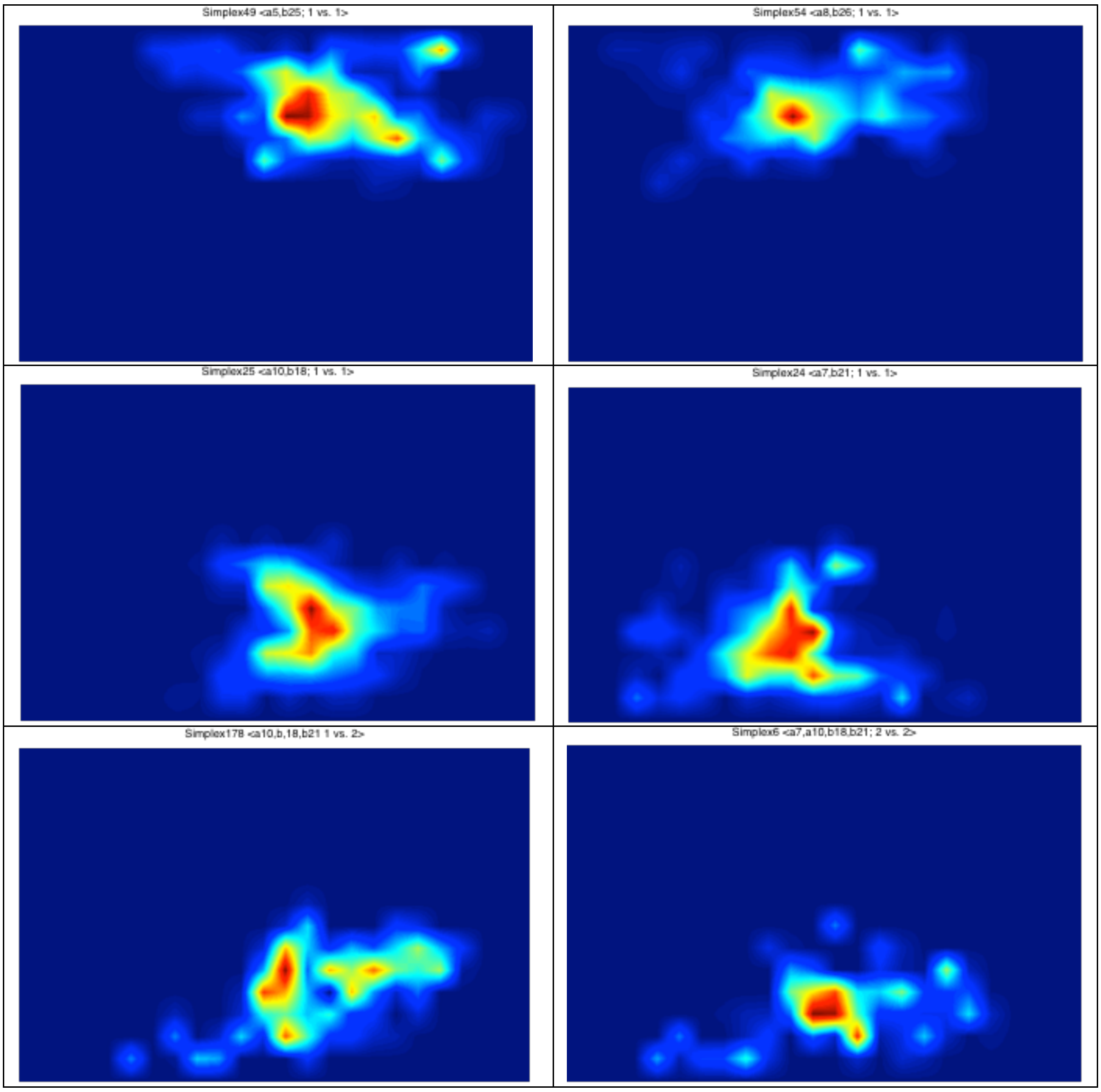


Figure 1. Heat map for simplices $\sigma_{49} = \langle a_5, b_{25}; (1\ vs. 1)\rangle$, $\sigma_{54} = \langle a_8, b_{26}; (1\ vs. 1)\rangle$ $\sigma_{25} = \langle a_{10}, b_{18}; (1\ vs. 1)\rangle$, $\sigma_{24} = \langle a_7, b_{21}; (1\ vs. 1)\rangle$, $\sigma_{178} = \langle a_{10}, b_{18}, b_{21}; (1\ vs. 2)\rangle$ and $\sigma_6 = \langle a_7, a_{10}, b_{18}, b_{21}; (2\ vs. 2)\rangle$.

In the meso level, we identified critical events dynamics such as:

i. Velocity of each player related to average velocity of the set;

ii. Changes of velocity and direction to break the symmetry of the set;

iii. Which players are central to break or maintain these symmetries.

The dynamics of simplices transformations near the goal depended on, significant changes in the players' speed and direction.

At macro level, we found how sets were related:

i. Emergent behavior analysis of players to promote local dominance analysis in critical events (see Figure 1);

Simplices are connected to one another, forming simplices of simplices including the goalkeeper and the goal.

Conclusions

The multilevel hypernetworks approach is a promising framework for soccer performance analysis once it captures cooperative and competitive interactions between players and sets of players. The spatiotemporal feature of the interactions between two or more players and sets of players are captured through the multilevel analyses and allows a richer understanding of real-world complex systems. Notably, players' moves can promote local dominance, i.e., moving to different directions from their closest players and increasing interpersonal distance; or moving to reduce interpersonal distances, either from their closest (typically) opponents or colleagues (local dominance). The identification of the most frequent simplices of players and their specific interactions, regarding local dominance, during a match is specific relevant information not only for analyzing the matches but also for preparation for future matches with different opponents.

References:

Araújo, D., & Davids, K. (2016). Team synergies in sport: Theory and measures. *Frontiers in Psychology, 7:1449. doi: 10.3389/fpsyg.2016.01449*.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., et al. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* 544, 1–122. doi: 10.1016/j.physrep.2014. 07.001.

Fonseca, S., Milho, J., Travassos, B., Araújo, D., and Lopes, A. (2013). Measuring spatial interaction behavior in team sports using superimposed Voronoi diagrams. *Int. J. Perform. Anal. Sport* 13, 179–189. doi: 10.1080/24748668.2013.11868640.

Headrick, J., Davids, K., Renshaw, I., Araújo, D., Passos, P., and Fernandes, O. (2012). Proximity-to-goal as a constraint on patterns of behaviour in attacker–defender dyads in team games. *J. Sports Sci.* 30, 247–253. doi: 10.1080/02640414.2011.640706.

Johnson, J. (2013). *Hypernetworks in the Science of Complex Systems,* Vol. 3. London: Imperial College Press.

Passos, P., Milho, J., Fonseca, S., Borges, J., Araújo, D., and Davids, K. (2011). Interpersonal distance regulates functional grouping tendencies of agents in team sports. *J. Mot. Behav.* 43, 155–163. doi: 10.1080/00222895.2011.552078.

4 Final remarks, perspectives and future work

Complexity sciences tools such as network science and graph theory have gained since the late 1990s considerable attention, due to their simplicity and communicative power and generality to be applied in different fields of science.

In sports settings, the use of network science and theory approaches began in this decade (since 2010s), mostly allied to PA as sub-science from sports sciences. PA were at this time (1980s) in search for a more multidisciplary approaches, beginning this path with DST and later (2010s) with network theory.

The main goal of this thesis was to explore empiricaly complex networks approach to team performance analysis, namely in team ball sports, like soccer. This sports context is rich in the availability of data and tecnhology, which elects it for most of the studys with network science approach.

Our innovative contribution is based on the evolution from static networks that were capturing only the dynamics on the networks to MHA where it is possible to represent the dynamics of the networks without overlooking the underlying structures and non-trivial topological patterns. Paper I is such an example, exposing some common pitfalls from the direct use of network metrics in different complex systems. In this review paper, the focus is on the reductionist contributs from SNA in PA, showing some possible directions to obtain the dynamics of the networks, mainly through bipartite and multilayer networks, where hypernetworks are an example.

The multilevel hypernetwork approach is based mainly on Johnson works, but according to papers in sections 3.2 and 3.3, with innovative extensions by: introducing compound variables (e.g. local dominance) that captures the structure and dynamics of cooperative and competitive interactions, the domain specificity of the soccer matches (solving mostly the "word problem"); including the spatiotemporal occurrence of the different sets of players (location and frequency); analyzing the dynamics in specific levels considering critical events and improving the details of dynamics in the mathematical formalism; and also exploring the dynamics of the sets transformations (e.g. interactions between simplices of simplices). We believe that the first question that is put forward in this thesis work (What are the structural and dynamical properties in cooperative (synergetic) and competitive interactions that most influence their performance outcomes?), can be clarified with these contributions from our work.

Considering sports as complex systems and specifically team sports, the potential for interaction-based situations increases, which in turn increases the complexity of the context in which performance is analysed. In our studies we have investigated some of the the most usual complexity features, like: a system with many heterogeneous parts; the dynamics of emergence from interactions of autonomous agents; unexpected or unpredictable emergence; multiple subsystem dependencies; self-organization into new structures and behaviors; adaptation to changing environments; co-evolving subsystems; multilevel dynamics; unrepeatable experiments; power-law regularities (Balague et al., 2013; Johnson, 2013; Juarrero, 2010; Komulainen, 2004; Piantadosi, 2014) and design in complex social systems (Alexiou et al., 2009; Blecic, 2008; Johnson, 2005b).

The search for a theoretical framework that could tackle more complexity related features lead us to use complex networks theories and tools, considering not only bipartite and temporal networks, but also multilayer networks capable of describing different levels of structure (from more stable ones to more dynamic ones) and *n*-ary cooperative and competitive interactions. This scalability property of multilevel hypernetworks, based on interpersonal distance (e.g. in section 3.2.2 and III, using the closest player criteria for the definition of simplices sets, which is a non-parametric variable) allowed to represent the instant of time in the mathematical formalism and both, the structure (backloth) and the dynamics (traffic) of the match, from a micro-level of analysis (e.g. typically the units/players of the system), to a meso-level (e.g. typically the structures/sets/teams) and a macro-level of analysis (e.g. the dynamics of any given event/GSO). The second proposed question, on this thesis work (Is the complex networks approach and its related tools able to identify, at different levels of analysis, the structure and the dynamics of the cooperative and competitive interactions in team sport complex systems, considering the results, the classifications and the match time?) had been, some how, answered at this point.

Through the analysis of the statistical distribution of the occurrence of the simplices in 10 soccer matches (from the 2010/2011 English Premier League Season), we found other interesting complexity features. These, are hallmark features present in other complex systems, like the scaling properties of human and natural phenomena usually described by power law type models. Population distribution in big cities, forest fires, forest patch sizes, scientific citations, WWW surfing, ecology, solar flares, economic index, epidemics in isolated populations, and the goal scoring distribution by players in soccer are some typical examples. In our study (section 3.3), we have found that the Zipf empirical law and its generalization by Mandelbrot was also present in the systems' meso-scale properties, like the processes that leads to a particular simplices' set occurrence distribution. Therefore it revealed common signatures of chaotic processes that are at one point self-organized and emerge from many natural and social processes. On the other hand, the tests to the validity of the ZM model, revealed some exceptions to these power laws. The few cases of simplices whose occurrence frequency are so high that they did not fit into ZM model, can be explained by the possible impact of Design. The deliberate design that expresses preformed intentions and purpose is revealed through some specific simplices. The match strategy designed by coaches and implemented through tactics by the players, has some distinct aspects from other complex social systems, once it also refers to competition interactions and not only cooperation interactions (Johnson, 2005a). This feature is clearly identified in the purpose of the narrow role of the goalkeepers, due to their connection/atraction to the Goals and the intentionality expressed by the symmetry positions of the players that are opponent (e.g. right defender from one team vs. the left attacker of the other team).

Our findings with multilevel hypernetworks approach were promising, not only from research perspetive but also considering practitioners perspetive. Presenting and discussing our results in different congresses/symposia and completing our multidisciplinary research team with a researcher/practitioner from the field we have some feedback that lead us to explore different variables and pointing out some limitations, like the absence of the ball positioning. Meanwhile, we were invited to participate in some emerging studies on hypernetworks, precisely introducing ball spatial positioning. The study from Ribeiro and colleagues (2019) where we contributed improving the formalisms of the simplices, specifically introducing ball position, the players who possed it and the actions they do with it, in the according simplices (Ribeiro et al., 2019). Ribeiro and colleagues are investigating the movement synchronization of the players whitin and between teams. In this study, the authors used MHA to access the interactions through player-simplex synchronies (Ribeiro et al., 2019). The third proposed question (Are the results obtained from complex network analysis useful in training/preparing situations? What are the structural properties that can help transmitting ecological validity from training to competition?), was in some way, also tackled, considering the two last paragraphs.

One of the challenges in our current work is to represent through the simplex mathematical formalisms, the positioning of the players inside the simplices (related to Goals). In the working progress of one of our studies in preparation (see reference VI, in Publications) we are proposing that the proximity to the Goals from each player could determinate the order of his representation inside the simplex formalism. In this new formalism, we are introducing an index value that represents the relationships inside the simplices. This proposal is based in the principle of inertia and uses as reference, the velocity of the geometrical center of the simplices and the simplex players' velocities, as for their instant contribute to aggregation or disaggregation of the simplex (see reference VI, in Publications). The fourth and last question proposed (What are the relevant types of interactions for the analysis of the structure and dynamics of the cooperative and competitive interactions between team players?), is starting to be answered in working progress that is still to come. However, some important ideas have been placed in order to tackle these problems.

Although we are very optimistic about the application and insights that can be provided by MHA in sports there is much ground yet to be explored, for example:

i. The impact (in PA) of other criteria for the simplices formation (MHA), like relative velocity, between simplices players' and also the ball;
ii. Analyzing simplices aggregation and disaggregation through temporal evolution algorithms, which allows to identify the dynamics and evolution of networks clustering.

References

A. Davis, B. B. G. a. M. R. G. Y. (1941). : Deep South: A Social Anthropological Study of Caste and Class Publisher: University of South Carolina Press Short Title: Deep South: A Social Anthropological Study of Caste and Class ISBN: 9781570038150.

Abreu, P., Moreira, J., Costa, I., Castelão, D., Reis, L., & Garganta, J. (2012). Human versus virtual robotics soccer: A technical analysis. *European Journal of Sport Science, 12*(1), 26-35.

Abreu, P. H., Moura, J., Silva, D. C., Reis, L. P., & Garganta, J. (2012). Performance analysis in soccer: a Cartesian coordinates based approach using RoboCup data. *Soft Computing, 16*(1), 47-61.

Albert-Laszlo, B., & Reka, A. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509-512.

Alexiou, K., Johnson, J., & Zamenopoulos, T. (2009). *Embracing complexity in design*: Routledge.

Almeida, F., Abreu, P. H., Lau, N., & Reis, L. P. (2012). *Automatic extraction of goal-scoring behaviors from soccer matches.* Paper presented at the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems.

Araujo, D., Davids, K., Bennett, S., Button, C., & Chapman, G. (2004). Emergence of sport skills under constraints

In Routledge (Ed.), *Skill Acquisition in Sport* (pp. 409-434): Taylor & Francis e-Library.

Araújo, D., & Bourbousson, J. (2016). Theoretical perspectives on interpersonal coordination for team behavior. *Interpersonal Coordination and Performance in Social Systems*, 126-139.

Araújo, D., & Davids, K. (2016). Team synergies in sport: Theory and measures. *Frontiers in Psychology, 7*.

Araújo, D., Davids, K., & Hristovski, R. (2006). The ecological dynamics of decision making in sport. *Psychology of Sport and Exercise, 7*(6), 653-676. doi:10.1016/j.psychsport.2006.07.002

Araújo, D., Lopes, R. J., Ramos, J., Silva, J., & Silva, C. (2014). EUA-Portugal à lupa: pouco ataque pelo centro e pouca precisão no remate. Retrieved from https://http://www.publico.pt/2014/06/23/desporto/noticia/euaportugal-a-lupa-pouco-ataque-pelo-meio-e-pouca-precisao-no-remate-1660182

Atkin, R. (1972). Multi-dimensional structure in the game of chess. *International Journal of Man-Machine Studies, 4*(3), 341-362.

Atkin, R. (1974). *Mathematical structure in human affairs*: Heinemann Educational Publishers.

Atkin, R., Hartston, W., & Witten, I. (1976). Fred CHAMP, positional-chess analyst. *International Journal of Man-Machine Studies, 8*(5), 517-529.

Atkin, R., & Witten, I. H. (1975). A multi-dimensional approach to positional chess. *International Journal of Man-Machine Studies, 7*(6), 727-750.

Balague, N., Torrents, C., Hristovski, R., Davids, K., & Araújo, D. (2013). Overview of complex systems in sport. *Journal of Systems Science and Complexity, 26*(1), 4-13. doi:10.1007/s11424-013-2285-0

Balkundi, P., & Kilduff, M. (2006). The ties that lead: A social network approach to leadership. *The Leadership Quarterly, 17*(4), 419-439. doi:10.1016/j.leaqua.2006.01.001

Bar-Yam, Y. (1997). *Dynamics of complex systems* (Vol. 213): Addison-Wesley Reading, MA.
Bar-Yam, Y. (2003). Complex systems and sports: complex systems insights to building effective teams. *Cambridge (MA): NECSI.*
Bar-Yam, Y. (2004). *Making things work: solving complex problems in a complex world*: Knowledge Industry.
Barrett, S., Genter, K., Hester, T., Quinlan, M., & Stone, P. (2010). *Controlled kicking under uncertainty.* Paper presented at the The Fifth Workshop on Humanoid Soccer Robots at Humanoids.
Bartlett, R., Button, C., Robins, M., Dutt-Mazumder, A., & Kennedy, G. (2012). Analysing Team Coordination Patterns from Player Movement Trajectories in Soccer: Methodological Considerations. *International Journal of Performance Analysis in Sport, 12*(2), 398-424.
Bartlett, R. M., Stockill, N. P., Elliott, B. C., & Burnett, A. F. (1996). The biomechanics of fast bowling in men's cricket: a review. *J Sports Sci, 14*(5), 403-424. doi:10.1080/02640419608727727
Beek, P. J., & Beek, W. J. (1988). Tools for constructing dynamical models of rhythmic movement. *Human Movement Science, 7*(2–4), 301-342. doi:http://dx.doi.org/10.1016/0167-9457(88)90015-2
Berkowitz, S. D., & Wellman, B. (1988). *Social structures : a network approach / edited by Barry Wellman, S.D. Berkowitz*. Cambridge [Cambridgeshire] ; New York :: Cambridge University Press.
Blecic, I. (2008). Design beyond complexity: Possible futures—Prediction or design?(and techniques and tools to make it possible). *Futures, 40*(6), 537-551.
Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., . . . Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports, 544*(1), 1-122.
Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: Structure and dynamics. *Physics Reports, 424*(4-5), 175-308. doi:10.1016/j.physrep.2005.10.009
Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science, 323*(5916), 892-895. doi:10.1126/science.1165821
Bourbousson, J., Seve, C., & McGarry, T. (2010). Space-time coordination dynamics in basketball: Part 2. The interaction between. *J Sports Sci, 28*(3), 349-358. doi:10.1080/02640410903503640
Brandes, U., Freeman, L., & Wagner, D. (2012). Social Networks. In R. T. (ed.) (Ed.), *Handbook of Graph Drawing and Visualization*. CRC Press.
Clemente, F. M., Couceiro, M. S., Martins, F. M. L., & Mendes, R. S. (2014). Using network metrics to investigate football team players' connections: A pilot study. *Motriz: Revista de Educação Física, 20*(3), 262-271.
Clemente, F. M., Couceiro, M. S., Martins, F. M. L., & Mendes, R. S. (2015). Using network metrics in soccer: A macro-analysis. *Journal of human kinetics, 45*(1), 123-134.
Clemente, F. M., Martins, F. M. L., Kalamaras, D., Wong, D. P., & Mendes, R. S. (2015). General network analysis of national soccer teams in FIFA World Cup 2014. *International Journal of Performance Analysis in Sport, 15*(1), 80-96.
Correia, V., Araujo, D., Craig, C., & Passos, P. (2011). Prospective information for pass decisional behavior in rugby union. *Human Movement Science, 30*(5), 984-997. doi:10.1016/j.humov.2010.07.008

Correia, V., Araújo, D., Duarte, R., Travassos, B., Passos, P., & Davids, K. (2012). Changes in practice task constraints shape decision-making behaviours of team games players. *Journal of Science and Medicine in Sport, 15*(3), 244-249. doi:10.1016/j.jsams.2011.10.004

Criado, R., Romance, M., & Vela-Pérez, M. (2010). Hyperstructures, a new approach to complex systems. *International Journal of Bifurcation and Chaos, 20*(03), 877-883.

Davids, K. (2008). *Dynamics of skill acquisition : a constraints-led approach / Keith Davids, Chris Button, Simon Bennett*. Champaign, IL :: Human Kinetics.

Davids, K., & Araújo, D. (2010). The concept of 'Organismic Asymmetry' in sport science. *Journal of Science and Medicine in Sport, 13*(6), 633-640. doi:10.1016/j.jsams.2010.05.002

Duarte, R., Araujo, D., Correia, V., & Davids, K. (2012). Sports teams as superorganisms: implications of sociobiological models of behaviour for research and practice in team sports performance analysis. *Sports Med, 42*(8), 633-642.

Duch, J., Waitzman, J. S., & Amaral, L. A. N. (2010). Quantifying the Performance of Individual Players in a Team Activity. *PLoS ONE, 5*(6), e10937. doi:10.1371/journal.pone.0010937

Dutt-Mazumder, A., Button, C., Robins, A., & Bartlett, R. (2011). Neural Network Modelling and Dynamical System Theory: Are They Relevant to Study the Governing Dynamics of Association Football Players? *Sports Med*. doi:10.2165/11593950-000000000-00000

Easley, D., & Kleinberg, J. Y. (2010). : Networks, Crowds, and Markets: Reasoning About a Highly Connected World Publisher: Cambridge University Press Number of Pages: 800 Short Title: Networks, Crowds, and Markets: Reasoning About a Highly Connected World ISBN: 0521195330, 9780521195331.

Erdös, P., & Rényi, A. (1959). On random graphs I. *Publ. Math. Debrecen, 6*, 290-297.

Fewell, J. H., Armbruster, D., Ingraham, J., Petersen, A., & Waters, J. S. (2012). Basketball Teams as Strategic Networks. *PLoS ONE, 7*(11), e47445.

Frencken, W., Lemmink, K., Delleman, N., & Visscher, C. (2011). Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science, 11*(4), 215-223. doi:10.1080/17461391.2010.499967

Girvan, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, 99*(12), 7821-7826. doi:10.1073/pnas.122653799

Glazier, P. S. (2010). Game, set and match? Substantive issues and future directions in performance analysis. *Sports Med, 40*(8), 625-634. doi:10.2165/11534970-000000000-00000

Glöckner, A., Heinen, T., Johnson, J. G., & Raab, M. (2012). Network approaches for expert decisions in sports. *Human Movement Science, 31*(2), 318-333.

Gould, P., & Gatrell, A. (1980). A structural analysis of a game: the Liverpool v Manchester United Cup Final of 1977. *Social Networks, 2*(3), 253-273.

Grund, T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks, 34*(4), 682-690. doi:10.1016/j.socnet.2012.08.004

Guillaume, J.-L., & Latapy, M. (2006). Bipartite graphs as models of complex networks. *Physica A: Statistical Mechanics and its Applications, 371*(2), 795-813. doi:http://dx.doi.org/10.1016/j.physa.2006.04.047

Haken, H. (1983). *Synergetics an introduction ; nonequilibrium phase transitions and self-organization in physics, chemistry, and biology* (3., rev. and enl. ed.). Berlin u.a.: Springer.

Haken, H., Kelso, J. A., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biol Cybern, 51*(5), 347-356.

Hughes, M., & Franks, I. M. (2008). *The essentials of performance analysis : an introduction*. London ; New York: Routledge.

Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences, 20*(10), 739-754. doi:10.1080/026404102320675602

Hölldobler, B., & Wilson, E. O. (2009). *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies*: Norton.

Johnson, J. (2005a). Complexity science in collaborative design. *CoDesign, 1*(4), 223-242.

Johnson, J. (2005b). *Multilevel Multidimensional Networks for Designing and Managing Complex Human Systems.* Paper presented at the Conference on Complexity Science and Society.

Johnson, J. (2006). Hypernetworks for reconstructing the dynamics of multilevel systems.

Johnson, J. (2008). Multidimensional Events in Multilevel Systems. In S. Albeverio, D. Andrey, P. Giordano, & A. Vancheri (Eds.), *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach* (pp. 311-334). Heidelberg: Physica-Verlag HD.

Johnson, J. (2013). *Hypernetworks in the science of complex systems* (Vol. 3): Imperial College Press London.

Johnson, J., Fortune, J., & Bromley, J. (2017). Systems, Networks, and Policy. In J. Johnson, A. Nowak, P. Ormerod, B. Rosewell, & Y.-C. Zhang (Eds.), *Non-Equilibrium Social Science and Policy: Introduction and Essays on New and Changing Paradigms in Socio-Economic Thinking* (pp. 111-134). Cham: Springer International Publishing.

Johnson, J. H. (2016). Hypernetworks: Multidimensional relationships in multilevel systems. *The European Physical Journal Special Topics, 225*(6), 1037-1052. doi:10.1140/epjst/e2016-02653-4

Johnson, J. H., & Iravani, P. (2007). The multilevel hypernetwork dynamics of complex systems of robot soccer agents. *ACM Transactions on Autonomous and Adaptive Systems (TAAS), 2*(2), 5.

Juarrero, A. (2010). Complex Dynamical Systems Theory. *Cognitive Edge.*

Katz, N., Lazer, D., Arrow, H., & Contractor, N. (2004). Network Theory and Small Groups. *Small Group Research, 35*(3), 307-332. doi:10.1177/1046496404264941

Kelso, J. A. (1984). Phase transitions and critical behavior in human bimanual coordination. *Am J Physiol, 246*(6 Pt 2), R1000-1004.

Kelso, J. A., & Jeka, J. J. (1992). Symmetry breaking dynamics of human multilimb coordination. *J Exp Psychol Hum Percept Perform, 18*(3), 645-668.

Kelso, J. A. S. (1995). *Dynamic patterns : the self-organization of brain and behavior*. Cambridge, Mass.: MIT Press.

Kelso, J. A. S., Buchanan, J. J., & Wallace, S. A. (1991). Order parameters for the neural organization of single, multijoint limb movement patterns. *Experimental Brain Research, 85*(2), 432-444. doi:10.1007/BF00229420

Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., & Osawa, E. (1997). *RoboCup: The Robot World Cup Initiative*. Paper presented at the Proceedings of the first international conference on Autonomous agents.

Komulainen, T. (2004). Self-similarity and power laws. *Complex systems-science on the edge of chaos (Report 145*.

Kugler, P., & Turvey, M. (1987). *Information, natural laws, and self-assembly of rhythmicmovement*: L. Erlbaum Associates (Hillsdale, N.J.).

Lees, A., & Nolan, L. (1998). The biomechanics of soccer: a review. *J Sports Sci, 16*(3), 211-234. doi:10.1080/026404198366740

Malacarne, L., & Mendes, R. (2000). Regularities in football goal distributions. *Physica A: Statistical Mechanics and Its Applications, 286*(1-2), 391-395.

Martinovic, J., Snásel, V., Ochodkova, E., Nolta, L., Wu, J., & Abraham, A. (2010). *Robot Soccer-Strategy Description And Game Analysis*. Paper presented at the ECMS.

McGarry, T. (2009). Applied and theoretical perspectives of performance analysis in sport: Scientific issues and challenges. *International Journal of Performance Analysis in Sport, 9*(1), 128-140.

McGarry, T., Anderson, D. I., Wallace, S. A., Hughes, M. D., & Franks, I. M. (2002). Sport competition as a dynamical self-organizing system. *J Sports Sci, 20*(10), 771-781. doi:10.1080/026404102320675620

Michaels, C., & Beek, P. (1995). The State of Ecological Psychology. *Ecological Psychology, 7*(4), 259-278. doi:10.1207/s15326969eco0704_2

Moody, J., McFarland, D., & Bender-Demoll, S. (2005). Dynamic network visualization. *American journal of sociology, 110*(4), 1206-1241.

Newman, M., Barabási, A.-L., & Watts, D. (2006). The Structure and Dynamics of Networks: (Princeton Studies in Complexity).

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review, 45*(2), 167-256.

Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks, 27*(1), 39-54. doi:http://dx.doi.org/10.1016/j.socnet.2004.11.009

Newman, M. E. J. (2010). *Networks : an introduction*. Oxford ; New York: Oxford University Press.

Passos, P., Araújo, D., & Davids, K. (2012). Self-Organization Processes in Field-Invasion Team Sports. *Sports Medicine*, 1-7. doi:10.1007/s40279-012-0001-1

Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., & Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport, 14*(2), 170-176. doi:10.1016/j.jsams.2010.10.459

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review, 21*(5), 1112-1130. doi:10.3758/s13423-014-0585-6

Ramos, J., Lopes, R. J., & Araújo, D. (2017). What's Next in Complex Networks? Capturing the Concept of Attacking Play in Invasive Team Sports. *Sports Medicine*. doi:10.1007/s40279-017-0786-z

Ramos, J., Lopes, R. J., Marques, P., & Araújo, D. (2017a). Hypernetworks Reveal Compound Variables That Capture Cooperative and Competitive Interactions in a Soccer Match. *Frontiers in Psychology, 8*, 1379. doi:10.3389/fpsyg.2017.01379

Ramos, J., Lopes, R. J., Marques, P., & Araújo, D. (2017b). *Hypernetworks: capturing the multilayers of cooperative and competitive interactions in soccer.* Paper presented at the International Congress Complex Systems in Sport.

Reilly, T., Cabri, J., & Araújo, D. (2005). Science and Football V. In. The proceedings of the world congress on science and football London: Routledge. (1st edition).

Ribeiro, J., Davids, K., Araújo, D., Silva, P., Ramos, J., Lopes, R., & Garganta, J. (2019). The Role of Hypernetworks as a Multilevel Methodology for Modelling and Understanding Dynamics of Team Sports Performance. *Sports Med*. doi:10.1007/s40279-019-01104-x

Riley, M., Richardson, M., Shockley, K., & Ramenzoni, V. (2011). Interpersonal Synergies. *Frontiers in Psychology, 2*(38). doi:10.3389/fpsyg.2011.00038

Rocha, L. (2011). *Exploring patterns of empirical networks*. UmeÃ¥: UmeÃ¥ University, Department of Physics.

Salingaros, N. A., & West, B. J. (1999). A universal rule for the distribution of sizes. *Environment and Planning B: planning and Design, 26*(6), 909-923.

Schmidt, R. C., Carello, C., & Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of. *J Exp Psychol Hum Percept Perform, 16*(2), 227-247.

Schmidt, R. C., O'Brien, B., & Sysko, R. (1999). Self-organization of between-persons cooperative tasks and possible application to sport. *International Journal of Sport Psychology, 30*(4), 558-579.

Solomonoff, R., & Rapoport, A. (1951). Connectivity of random nets. *The bulletin of mathematical biophysics, 13*(2), 107-117.

Travassos, B., Bourbousson, J., Esteves, P. T., Marcelino, R., Pacheco, M., & Davids, K. (2016). Adaptive behaviours of attacking futsal teams to opposition defensive formations. *Human Movement Science, 47*, 98-105. doi:http://dx.doi.org/10.1016/j.humov.2016.02.004

Travers, J., & Milgram, S. (1977). An experimental study of the small world problem. In *Social Networks* (pp. 179-197): Elsevier.

Vilar, L., Araújo, D., Davids, K., & Button, C. (2012). The role of ecological dynamics in analysing performance in team sports. *Sports Med, 42*(1), 1-10. doi:10.2165/11596520-000000000-00000

Watts, D. (2003). : Six Degrees: The New Science Publisher: Random House Short Title: Six Degrees: The New Science ISBN: 9780099444961.

Williams, A. M., & Hodges, N. J. (2004). *Skill acquisition in sport : research, theory and practice*. London ; New York: Routledge.

Zhu, G., Xu, C., Huang, Q., Rui, Y., Jiang, S., Gao, W., & Yao, H. (2009). Event tactic analysis based on broadcast sports video. *IEEE Transactions on Multimedia, 11*(1), 49-67.

Zipf, G. K. (1949). Human behavior and the principle of least effort.