

CASE STUDY

Open Access



The network of sports: using network analysis to understand the relationship between sports and socio-physiological factors in contemporary China

Xiangyang Bi¹ , Zhanning Sun^{2*} and Boran Hu³

*Correspondence:
sun_zhanning@126.com

¹ Minzu University of China,
Beijing, China

² China Sports Policy Research
Institute, Beijing Sport University,
Beijing 100084, China

³ China Institute of Sport Science,
Beijing, China

Abstract

This study examines sports and physical activities among Chinese aged 18–65, using network analysis on a significant random sample. It categorizes sports into 11 groups based on public selection, with a community detection algorithm aiding classification. Variables like age, gender, and education were integrated, revealing how life stages, gender disparities, and social class influence activity participation. The mixed graph model shows both positive and negative correlations among activities, highlighting the role of sports as both a social integrator and divider, reflective of broader societal norms and inequalities. The exponential random graph model further illustrates a complex network of demographic-driven participation patterns. The purpose of this investigation is twofold: to advance methodological approaches in the study of sports-related social networks and to explore the broader implications such networks may have on individual and collective behaviors within this field.

Keywords: Sports and physical activities, Network analysis, Social distinction, Mixed graphical model, Sociology of sports

The questions I shall raise come from outside; they are the questions of a sociologist who, among the objects he studies, encounters sporting activities and entertainments in the form, for example, of the statistical distribution of sports activities by educational level, age, sex, and occupation, and who is led to ask himself questions not only about the relationship between the practices and the variables, but also about the meaning which the practices take on in those relationships.

—Pierre Bourdieu 1978

Background and issues

Acknowledging the rise of social network analysis and its previous marginalization in the realm of sports research, Nixon (1993, 2002) championed the integration of this analytical approach into the domain of sports studies. He suggested that this new approach helps sport sociologists look beyond individual traits to understand relationships and

complex social interactions in sports, revealing important dynamics within social networks (Nixon 1993, 2002). Heeding this call to action, the field has observed a discernible transformation. While still in its developmental stages, network analysis is forging a novel paradigm in sports science research. This is substantiated by an expanding corpus of literature that employs these methods for exploring subjects ranging from organizational behavior and performance benchmarks in sports (Onody and DeCastro 2004; Breznik and Batagelj 2012; Clemente and Rui 2016; Ribeiro et al. 2017), to dynamics of sports communication (Hambrick 2013; Hambrick and Marion 2017), as well as the investigation of social capital and support mechanisms (Norris et al. 2020; Gemar 2021), and beyond.

Considering the methodologies of network analysis, Wäsche et al. (2017) categorized relevant studies into distinct types: competitive networks, interactional networks, both intra- and extra-organizational networks, affiliation networks, and networks pertaining to the social milieu enmeshing sporting entities. In his latest scholarly work, Hambrick (2019) encapsulated the utilization of social network analysis across various tiers of sports entities, encompassing individuals, teams, organizations, and media, as well as within the broader scope of sports research. Nevertheless, when it comes to examining the nexus between sports and physical activities in and of themselves, the sociology of sport has yet to produce research employing network analysis on this particular subject.

Collectively, sports constitute a multifaceted system composed of numerous distinct events, necessitating a nuanced exploration of the relationship between sports and physical activities from diversified vantage points. Employing the network analysis methodology, this paper delves into the categorization of various social groups engaged in sports and physical activities through a lens informed by the sociology of sports. It scrutinizes the intricate interplay between sports and physical activities and examines the social determinants underpinning these relational patterns.

The intrinsic characteristics and socio-physiological factors of participants significantly affect the patterns of clustering that arise within different sports and physical activities. Such clustering unveils complex interrelations that challenge rudimentary analytical approaches premised on the independence and irrelevance of choices. Thereby, conventional analyses that consider variables in isolation fall short, for they fail to recognize these multifaceted interdependencies. In response to these methodological limitations, the current investigation harnesses the capabilities of mixed graphical model (MGM) and exponential random graph model (ERGM), underpinned by the descriptive statistics derived from survey data. MGM is proficient in delineating conditional dependencies among variables, thereby offering a refined perspective on the interplay between various activities. Complementarily, ERGM provides insights into the structural connections within these networks, highlighting the roles of variables such as gender, age, and social status (Lusher et al. 2013).

Literature review and research hypothesis

Firstly, initial examinations of extant literature (Gidlow et al. 2006; Kahma 2012; Federico et al. 2013) reveal a discernible link between social class and engagement in sports and physical activities, with higher social strata more frequently participating in activities of an above-moderate intensity level. Although sports may transcend

social boundaries to a degree, there remains a stark variance in how different social classes partake in various sports and physical activities. Bourdieu (1978) accentuated the contrast between the elite-oriented sports such as tennis, equestrianism, sailing, and golf, and those more commonly associated with the lower-middle class, such as basketball, rugby, cycling, football, boxing, and wrestling—each reflective of distinct social spheres. These preferences and proclivities, deemed as class habitus, are rooted in a system influenced by a myriad of factors, including economic capital—which encompasses leisure time—cultural capital, and the ethical and aesthetic predilections of distinct social classes. Fussell (1992: 112) referenced Birnbach's observation that the upper echelons of society tended to favor sports involving smaller balls, which was indicative of underlying class differences. Consequently, this research posits Hypothesis 1 (H1): There is a significant divergence in the participation within different sports and physical activities along the social class spectrum.

Secondly, there exists a notable correlation between the life course—specifically, age—and engagement in sports and physical activities, which is inherently linked to socio-physiological considerations. This manifests as an age cohort effect in physical activities, with younger cohorts demonstrating more frequent participation compared to their older counterparts (Hirvensalo and Lintunen 2011; Apostolou 2015; Gantz and Lewis 2021). Longitudinal research (Vandervoort et al. 2012; Aggio et al. 2018) has observed the continuity of participation in certain sports (e.g., golf) throughout the lifespan but has often overlooked the shifts in preferences for specific sports with advancing age. Despite the straightforward nature of this observation, there remains a conspicuous scarcity of systematic, quantitative investigations into the selection of specific physical activities by different age demographics. With a foundational comprehension of this subject, this paper introduces Research Hypothesis 2 (H2): There is considerable variability in participation within different sports and physical activities across various age groups.

Thirdly, gender segregation exists in sports and physical activities, particularly in certain sports characterized by prominent gender stereotypes (Koivula 2001; Hardin and Greer 2009; Chinurum et al. 2014; Bandy 2014). Notwithstanding, scholarly inquiries have predominantly examined professional sports contexts, paying less attention to gender disparities in the realm of communal sports engagement. It has been recognized (Paloian 2012) that traditionally masculine sports such as boxing, ice hockey, weightlifting, and motorsports contrast with those activities where female participation is typically encouraged, including figure skating, gymnastics, and tennis. In light of these observations, this paper posits Research Hypothesis 3 (H3): There exists a significant gender-based differentiation in the participation in sports and physical activities.

Examining the available literature reveals that conducting a detailed and comprehensive review of the distribution patterns of various sports and their connections to societal impacts remains a significant challenge. This difficulty is partly due to a reality where a vast number of studies are primarily focused on secondary analyses of generic social survey data. Although some research endeavors are thematically focused, they typically concentrate on a restricted set of sports or engage with the concept of sports participation in a broad sense, rather than conducting an exhaustive study of specific sports and physical activities. Utilizing national data from a dedicated survey on sports and physical

exercises, the current research offers an opportunity for a more detailed examination of this subject area.

Methodology

Survey and data

To fulfill the assignment from the China Basketball Association, our research team engaged a specialized cloud-based survey platform for conducting an online panel survey concerning general sports and physical activities.¹ Drawing upon the 2015' 1% Population Sample Statistical Yearbook issued by the National Bureau of Statistics, this project utilized a multi-stage, unequal probability stratified random sampling method, stratifying by region, urban–rural classification, and age across eight geographic areas of China (excluding Hong Kong, Macao, and Taiwan). The survey, employing a web-based questionnaire, reached participants through an online sampling pool from January 21 to February 3, 2021. Of the 7081 questionnaires received, 7075 met the quality standards upon review. The analysis excluded responses indicating no participation in sports or physical activities (12.0% of the population aged 18–65 years) due to its logical exclusivity and disregarded the “other” option (chosen by 0.7% of the population aged 18–65 years) due to its minimal selection rate. Consequently, the final valid sample size for this study was determined to be 6464.

The design of the questionnaire was a collaborative effort by the project team. Data pertinent to this study were gleaned from responses to the question, “What sports and physical activities do you usually engage in?” included in the questionnaire. This question allowed for multiple responses from a list of 45 options (including “not doing any sports and physical activities” and “other” options). Respondents had the liberty to select options that best reflected their personal activities. To ensure the survey’s impartiality, its administration was entrusted to an independent third-party organization that conducts cloud surveys. The survey, under the title “Research on Public Attitudes and Behaviors towards Daily Activities in China,” was administered on behalf of the “China Sports Policy Research Institute, Beijing Sport University.” It is noteworthy that the questionnaire was generalized and did not contain inquiries specifically oriented towards basketball. In sections where participants and their offspring were queried about their “usual sports and physical activities” and “the most significant sport or physical activity” (the latter requiring a singular choice from the previously selected options), basketball was merely one among the 45 possibilities. This design ensured that the survey would not be unfairly biased towards basketball or against other types of sports due to the inclusion of basketball-specific questions. Additionally, to mitigate any order effect bias, the presentation of 45 options was randomized.

Following the collection of data, the data from the online sample survey were assessed across several demographics—gender, age, education level, region, and urban/rural classification—as delineated in the national 1% population sample survey. The dataset was then weighted through a combined approach of post hoc stratified weighting and raking

¹ The URL of the questionnaire is: <https://www.epanel.cn/evt/C0410B25A7F04B51DFBDE70400495078>. The preview version of pilot survey can be visited from: <https://www.epanel.cn/works/257429/preview>. The exclusive sample size of online panel pool is 2 million.

Table 1 Statistical description

| Variable | Mean | Std |
|--------------------|------------|--------|
| Years-of-schooling | 12.1 | 2.8 |
| Age | 37.8 | 12.3 |
| Gender | Percentage | |
| | Male | Female |
| | 47.9% | 52.1% |

weighting across all these demographic fields. The outcomes revealed that, after adjustment through weighting, the composition of the sample closely mirrored the overall demographic distribution of the census population, thereby endowing the sample with robust representativeness. This alignment allows for reliable extrapolations regarding the sports and physical activities engaged in by the national population aged 6–65.² Unless specifically indicated otherwise, the conclusions drawn in this study are all based on data subjected to this weighted analysis.

Description of background variables

Given the simplistic nature of the occupational metric within the questionnaire, this study adopts educational level as a proxy for the social class dimension. Educational level, which is intricately linked to tangible aspects such as occupation and income, serves as a principal indicator of an individual's cultural and social capital. In the context of this survey, the variable of educational level is recalibrated into years-of-schooling, corresponding to the respondents' educational phases; for instance, completion of elementary school is equated to 6 years, while junior high school completion is equated to 9 years. This re-coding renders educational level as a continuous variable. The average educational attainment among respondents is 12.1 years, with a standard deviation of 2.8 years. Within the sample demographic, males constitute 52.1%, while females represent 47.9%. The average age of respondents is 37.8 years, with a standard deviation of 12.3 years (see Table 1).

Research findings

Descriptive analysis

The survey findings indicate that among individuals aged 18–65 who engage in sports and physical activities, activities such as running, walking, and rope skipping are notably prevalent. In addition to these general physical activities, sports that are more accessible in terms of equipment and venue, including badminton, table tennis, and basketball, emerge as the most favored options (See Table 2).

Mixed graphical model

To concurrently ascertain the intricate interplay among various sports and physical activities and socio-physiological determinants such as age, gender, and educational

² The survey inquired about the athletic habits of the respondents, as well as those of their offspring. Of the respondents, 1928 had children ages 6–17 years. Nevertheless, owing to significant disparities in sports involvement between youth and adults, and the complexities associated with assessing educational levels in adolescents, this study focuses exclusively on the adult demographic.

Table 2 Selected sports and physical activity preferences and background variables in 18–65 year olds

| | | All | Physically active population | Percent of male | Percent of female | Mean of age | Mean of Edu |
|----|---|------|------------------------------|-----------------|-------------------|-------------|-------------|
| 1 | Football | 3.8 | 4.1 | 6.1 | 2.0 | 33.1 | 14.0 |
| 2 | Basketball | 9.5 | 9.2 | 14.3 | 3.7 | 31.7 | 13.7 |
| 3 | Volleyball | 2.0 | 1.9 | 2.6 | 1.2 | 33.6 | 14.5 |
| 4 | Badminton | 18.2 | 18.8 | 16.2 | 21.6 | 36.9 | 12.9 |
| 5 | Tennis | 2.4 | 2.5 | 2.8 | 2.2 | 34.8 | 14.3 |
| 6 | Table Tennis | 11.1 | 11.9 | 14.2 | 9.5 | 38.6 | 13.2 |
| 7 | e-Sports | 3.8 | 3.1 | 4.4 | 1.7 | 27.5 | 14.0 |
| 8 | Chess and cards | 9.6 | 10.5 | 11.7 | 9.1 | 44.7 | 12.0 |
| 9 | Run | 36.7 | 37.9 | 37.4 | 38.5 | 38.1 | 12.5 |
| 10 | Walk/Skelp/Speed walking | 21.5 | 21.3 | 19.5 | 23.3 | 42.6 | 11.7 |
| 11 | Equipment fitness/Bodybuilding | 3.3 | 3.3 | 2.5 | 4.1 | 39.6 | 13.1 |
| 12 | Community sports equipment activities | 9.4 | 10.2 | 10.4 | 10.0 | 45.6 | 11.8 |
| 13 | Golf | 1.3 | 1.2 | 1.7 | 0.6 | 37.7 | 14.4 |
| 14 | Gymnastics/Eurhythms | 1.5 | 1.4 | 1.0 | 1.9 | 37.1 | 12.6 |
| 15 | Aerobics/Body dance | 3.9 | 3.6 | 0.9 | 6.6 | 37.8 | 12.2 |
| 16 | Martial arts/Sanda/Wrestling | 1.2 | 1.3 | 2.3 | 0.3 | 35.5 | 13.6 |
| 17 | Swimming/Synchronized swimming | 2.7 | 3 | 4.0 | 1.9 | 36.6 | 14.0 |
| 18 | Yoga | 6.3 | 6.6 | 1.1 | 12.5 | 34.6 | 13.1 |
| 19 | Mountain-climbing/Rock climbing | 4.1 | 4.7 | 5.6 | 3.8 | 35.5 | 12.6 |
| 20 | Taekwondo/Judo/Karate | 1.5 | 1.2 | 1.6 | 0.8 | 33.6 | 14.7 |
| 21 | Cycling/Bicycle riding | 14.9 | 14.9 | 14.8 | 15.0 | 37.5 | 12.4 |
| 22 | Rope skipping | 11.0 | 10 | 6.1 | 14.1 | 32.8 | 12.8 |
| 23 | Roller-skating/Skateboard | 2.0 | 1.3 | 1.6 | 1.1 | 30.0 | 14.4 |
| 24 | Croquet | 1.3 | 1 | 1.2 | 0.7 | 40.9 | 13.3 |
| 25 | Boxing | 1.5 | 1.6 | 2.5 | 0.6 | 31.6 | 14.0 |
| 26 | Equestrian/Polo | 0.7 | 0.7 | 0.9 | 0.4 | 34.5 | 14.7 |
| 27 | Fencing | 1 | 1.0 | 1.5 | 0.5 | 38.7 | 14.7 |
| 28 | Bowling | 1.6 | 1.7 | 2.2 | 1.0 | 37.9 | 14.1 |
| 29 | Shooting/Archery | 1.3 | 1.3 | 1.8 | 0.7 | 34.5 | 13.4 |
| 30 | Skating/Skiing | 1.7 | 1.8 | 1.9 | 1.6 | 29.9 | 13.9 |
| 31 | Curling/Floor curling | 0.6 | 0.4 | 0.6 | 0.2 | 26.4 | 15.3 |
| 32 | Ice hockey | 0.8 | 0.8 | 1.2 | 0.4 | 30.8 | 14.6 |
| 33 | High jump/Long jump | 3.9 | 3.1 | 4.1 | 2.0 | 31.4 | 13.5 |
| 34 | Shot-put/Discus/Hammer-throwing/Javelin/Medicine ball | 1.0 | 1.0 | 1.5 | 0.4 | 34.3 | 14.0 |
| 35 | Outdoors/Directional off-load/Hiking | 10.6 | 11.1 | 13.8 | 8.2 | 38.3 | 12.6 |
| 36 | Billiards | 5.7 | 6.5 | 10.1 | 2.6 | 31.0 | 13.2 |
| 37 | Aquatics (Sailboat, surfing, diving etc.) | 1.9 | 1.5 | 1.8 | 1.1 | 29.5 | 14.6 |

Table 2 (continued)

| | All | Physically active population | Percent of male | Percent of female | Mean of age | Mean of Edu |
|----|--|------------------------------|-----------------|-------------------|-------------|-------------|
| 38 | Folk Traditional sports (Kick key, kite, diabolo etc.) | 6.0 5.6 | 5.9 | 5.3 | 39.9 | 12.4 |
| 39 | Hockey | 1.2 1.2 | 1.7 | 0.6 | 31.8 | 14.7 |
| 40 | Rugby | 1.7 1.5 | 2.2 | 0.8 | 33.5 | 13.8 |
| 41 | Model airplane/Model car/Radio direction finding | 1.2 1.2 | 1.3 | 1.0 | 30.9 | 13.2 |
| 42 | Racing | 1.7 1.8 | 2.5 | 1.0 | 30.9 | 13.5 |
| 43 | Pull-up/Sit-up/Sit and reach | 11.4 10.6 | 13.6 | 7.4 | 32.0 | 13.1 |
| 44 | Other | 0.7 – | – | – | – | – |
| 45 | No sports or exercise | 12.0 – | – | – | – | – |
| | Total | 251 239 | – | – | – | – |

Respondents' selections of sports and physical activities were made through a multiple-choice format, with the findings displayed as a percentage of cases

attainment, the present study utilizes an undirected probabilistic graphical model. The data encompasses binary variables for sports and physical activity options as well as gender, alongside continuous variables for age and educational attainment, suggesting an integration of Ising and Gaussian distributions. Thus, a mixed graphical model is employed for analytical purposes. Within this model, a simple undirected graph, $G = (V, E)$, is formulated, where V represents the set of nodes corresponding to attribute and continuous variables, and E denotes the set of edges. Each node symbolizes a particular variable, and the absence of an edge between any two nodes within the undirected graph signifies their conditional independence when all other variables are accounted for. Conversely, the presence of an edge indicates a conditional dependency (refer to the Fig. 1 for the model's specific values). In the graphical representation generated by the mixed graphical model's `mgm` function, various types of relationships are color-coded for clarity: green edges represent positive correlations, red edges signify negative correlations, and gray edges are used when the relationship is undefined or unspecified.

The mixed graphical model employed in this study leverages L1-Regularized (lasso) neighborhood regression, which is a technique grounded in machine learning algorithms designed to refine the model by controlling for exogenous edge parameters. This is achieved by optimizing the penalized conditional likelihood at each node, thereby providing a reliable and precise estimation of the network structure among the variables (Yang et al. 2014; Lee and Hastie 2015; Baker 2017; Haslbeck and Waldorp 2018). The deployment of this analysis was facilitated through the `mgm` package within the R programming environment (Haslbeck and Waldorp 2020).³

³ We use `msm` function to estimate k-degree mixed graphical model via nodewise regression. Here we are interested in fitting a pairwise MGM, and we, therefore, choose $k=2$. Tuning parameter takes cross-validation (`lambdaSel = "CV"`, `lambdaFolds = 10`). In default setting of `mgm` function, gaussian nodes are centered and divided by their standard deviation, i.e., `scale = TRUE`.

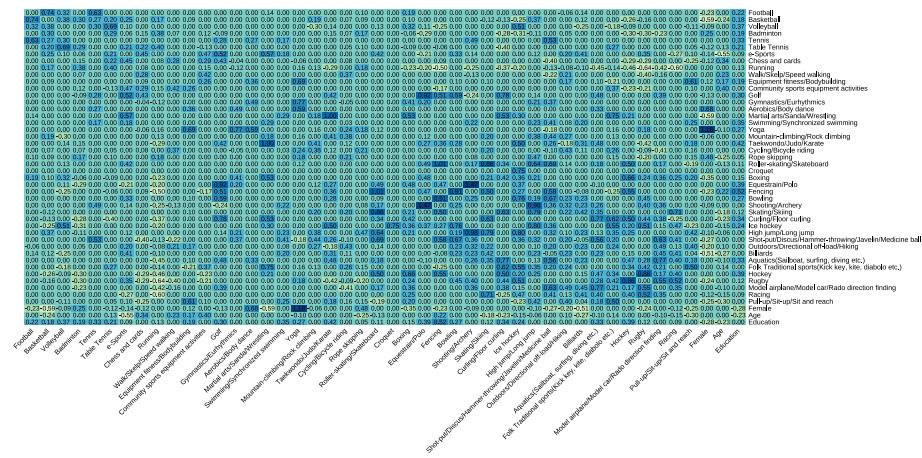


Fig. 1 Network coefficients matrix outputted by mixed graph model

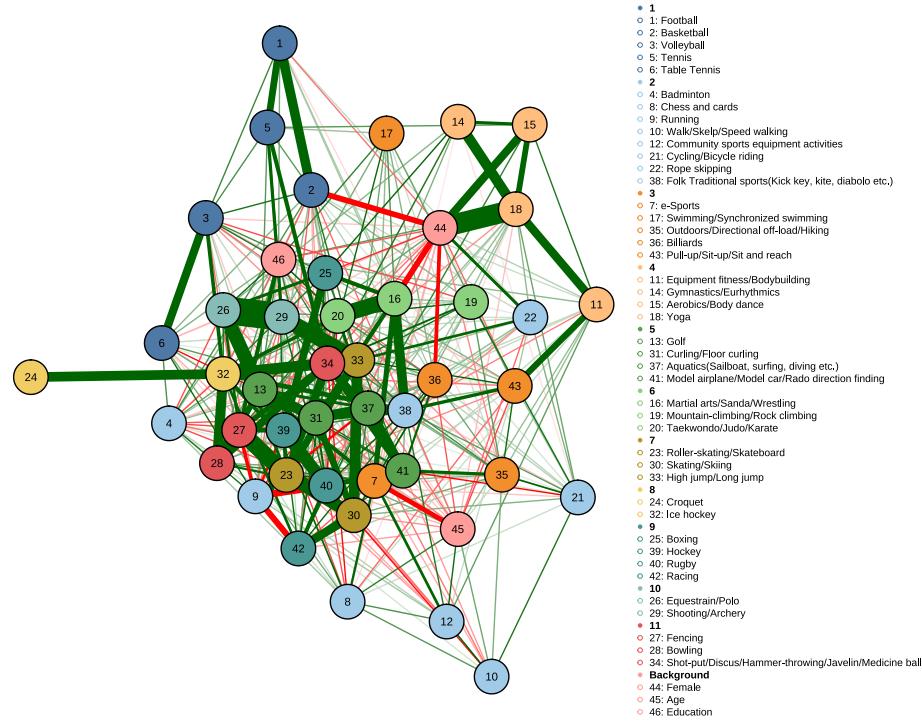


Fig. 2 Network of sports and physical activities based on MGM estimation

Employing the output matrix derived from the mixed graphical model, in the absence of background variables, the study proceeds with community detection using the spin-glass algorithm.⁴ This analytical approach successfully segments sports and physical activities into 11 distinct categories, as illustrated in Fig. 2.

⁴ Related options settings: implementation = 'neg', gamma = 2, spin = 20, weights = E(gra)\$weight.

- Category 1 includes Football (1), Basketball (2), Volleyball (3), Tennis (5) and Table Tennis (6), which are not only popular but also require public facilities. These activities are not only widely embraced by the public but also necessitate the utilization of communal facilities, underscoring the importance of community involvement and the imperative for infrastructural provision.
- Category 2 is characterized by a mass appeal and accessibility, featuring low-entry barriers and minimal physical demands. This category includes Badminton (4), Chess and Cards (8), Running (9), Walking/Speed Walking (10), Community Sports Equipment Activities (12), Cycling (21), Rope Skipping (22), and traditional folk sports (38) such as Kicking Key, Kite Flying, and Diabolo. Many of these activities are leisurely by nature and are commonly engaged in within community settings.
- Category 3 is diversified, comprising e-Sports (7), Swimming and Synchronized Swimming (17), Outdoors/Directional off-load/Hiking (35), Billiards (36), and exercises like Pull-ups, Sit-ups, and the Sit and Reach test (43). While some are predominantly youth-oriented leisure activities or components of physical education assessments, others are favored by young adults frequenting gaming arcades and similar venues.
- Category 4 is associated with body sculpting and is notably more prevalent among females. It includes activities such as Equipment Fitness/Bodybuilding (11), Gymnastics/Eurhythms (14), Aerobics/Dance Fitness (15), and Yoga (18).
- Category 5 is distinguished by its niche-oriented, elite nature, comprising sports such as Golf (13), Curling (31), various Aquatic activities including Sailing, Surfing, and Diving (37), and Model Aviation/Car Racing and Radio Direction Finding (41).
- Category 6 is geared toward combat and strength, appealing predominantly to male participants. It includes Martial Arts/Sanda/Wrestling (16), Mountain/Rock Climbing (19), and Taekwondo/Judo/Karate (20).
- Category 7 comprises activities like Roller-skating/Skateboarding (23), Ice Skating/Skiing (30), and High Jump/Long Jump (33). These sports have garnered popularity among urban youth, necessitating considerable physical flexibility and balance.
- Category 8 encompasses games such as Croquet (24) and Ice Hockey (32), sharing attributes of being goal-oriented and relatively niche-oriented, with a collective aspect to the sporting experience.
- Category 9 includes physically intensive sports like Boxing (25), Field Hockey (39), Rugby (40), and Racing (42). These have gained traction in urban regions of China in recent years, with most entailing significant physical contact and a higher susceptibility to injuries.
- Category 10 is represented by Equestrian/Polo (26) and Shooting/Archery (29). This category is indicative of an exclusive and niche-oriented segment of sports that have found a particular resonance among urban youth from affluent backgrounds.
- Category 11 features sports such as Fencing (27), Bowling (28), and various field events including Shot Put, Discus, Hammer Throw, Javelin, and Medicine Ball (34). These activities typically occur in specialized facilities or arenas and necessitate the use of protective professional equipment.

Overall, the correlation between sports and physical activities reveals a strong positive linkage among niche-oriented sports and a relatively weaker connection among mainstream sports. This trend persists in the interactions between mainstream and niche sports as they typically exhibit a more remote relationship. Specifically, a notable negative correlation is identified between certain sports and activities, exemplified by Tennis (5) and a range of traditional folk sports such as Kick Key, Kite Flying, and Diabolo (38). The data further indicates that Running (9) occupies a distinct niche within the realm of specialized sports, commanding widespread appeal among its adherents. These enthusiasts typically exhibit a negative correlation with engagement in other niche sporting activities. Such a pattern may stem from the considerable investment of time, energy, and resources that running—a strenuous endeavor demanding both physical and mental exertion—necessitates. Consequently, those devoted to the discipline of running prefer to focus their efforts on improving their capabilities within this singular pursuit, rather than diluting their focus by engaging in a variety of sports. Yet, positive correlations are observed between Golf (13) and Equestrian/Polo (26), as well as Fencing (27) and Bowling (28).

However, exceptions to these general trends exist. For instance, Football (1) and Basketball (2) demonstrate a significant correlation (0.74), indicating a considerable overlap among individuals who participate in both sports. This relationship is not one of mutual exclusion but rather one of mutual reinforcement. Furthermore, there is a modest positive correlation (0.25) between Basketball (2) and eSports (7). The data analysis regarding adolescents indicates that a pronounced triangular connection exists between Football (1), Basketball (2), and eSports (7) among participants aged 6 to 17, suggesting that eSports serves as a complement to, rather than a replacement for, traditional sports such as football and basketball.

Within the context of demographic variables, the female group demonstrates a pronounced positive correlation with activities such as Yoga (18), Aerobics/Body Dance (15), Rope Skipping (22), and Badminton (4). Excluding Equipment Fitness/Bodybuilding (11), these activities predominantly attract female participants in Category 4. Conversely, this group exhibits significant negative correlations with sports traditionally dominated by males, including Football (1),⁵ Basketball (2), Martial Arts/Sanda/Wrestling (16), Boxing (25), and field events like Shot Put, Discus, Hammer Throwing, Javelin, and Medicine Ball (34). Additionally, female participation is negatively correlated with Billiards (36) and strength training exercises such as Pull-ups, Sit-ups, and the Sit and Reach (43).

Age appears to have a notable negative correlation with engagement in eSports (7), Basketball (2), Rope Skipping (22), Roller-Skating/Skateboarding (23), and exercises such as Pull-ups, Sit-ups, and the Sit and Reach (43), among others. The likelihood of involvement in these sports and physical activities decreases with advancing age. Conversely, this trend is inverted for activities such as Chess and Card Games (8), Community Sports Equipment Activities (12), and Walking activities, including Skelp and Speed Walking (10), where participation tends to increase as age progresses.

⁵ Football enjoys greater popularity among females compared to basketball, partly attributable to the recent promotion of school-level football as a state policy in mainland China.

Within this analytical framework, an individual's level of educational attainment does not exhibit negative correlations with engagement in any sport or physical activity; instead, relationships are either neutral or positive. Specifically, individuals with higher educational qualifications demonstrate significant positive correlations with a variety of sports and physical activities. These include Fencing (27), Hockey (39), Tennis (5), Martial Arts such as Taekwondo, Judo, and Karate (20), aquatic activities like Swimming and Synchronized Swimming (17), Curling and Floor Curling (31), Golf (13), Volleyball (3), Yoga (18), Bowling (28), Ice Hockey (32), and other Aquatic sports (e.g., Sailing, Surfing, Diving) (37), among others.

Exponential random graph model

To delve deeper into the structural associations between various categories of sports and physical activities, as well as the impact of socio-physical characteristics, we employ an exponential random graph model (ERGM) (Harris 2013; Lusher et al. 2013). The ERGM articulates the likelihood of network ties (binary relationships) through a dual-component approach: the endogenous structural embeddings and the exogenous attributes of nodes. Integrating the endogenous configuration factor within the model circumvents the limitations inherent in conventional regression techniques, which often breach the assumption of non-independence.

Prior to conducting the modeling analysis, the output matrix derived from the mixed graph model (refer to Fig. 1) is processed such that elements with values exceeding 0 are recalibrated to 1, while those with values of 0 or below are recalibrated to 0. Subsequently, demographic variables including gender, age, and educational attainment are compiled by sport category to summarize overall characteristics. These aggregates, alongside data on sports clusters and participation rates, are then assigned to the attributes of the respective sports network nodes (detailed results can be found in Table 2). At the project level, all continuous variables—namely, the proportion of males, proportion of females, years-of-schooling, age, and the multiple responses associated with each sport—are normalized, i.e., each is centered and divided by its standard deviation. This normalization serves to mitigate the issue of comparability that arises from the inherent variation in the natural distribution of different factors across sports, thus enhancing the robustness of the model estimates. For the actual analysis, the “ergm” package in the R programming language is utilized to conduct the ERGM analysis (Hunter et al. 2008), with Monte Carlo maximum likelihood estimation (MCMLE) being applied to estimate model parameters.⁶ The findings are presented in Table 3.

The edge values in our analysis represent the mean influence that edges exert in the formation of networks, serving a function analogous to the intercept in ordinary least squares (OLS) regression models. According to our findings, the network model's density, considering only the edges, is calculated as 0.283, derived from the formula $[\exp(-0.927)/(1 + \exp(-0.927))]$. Within the realm of endogenous effects, the

⁶ Related options settings: `burnin = 15,000`, `MCMCsamplesize = 30,000`. Furthermore, employing the latent space model while excluding the aforementioned categorical factors from the explanatory variables yields results that are comparable to those of Model 4. In addition, inspection of the four subgraphs included in Fig. 3 demonstrates that the data generated by the current ERGM model exhibits a good fit with the empirical observations across all assessed indicators. This congruence indicates that the ERGM model in question is well-calibrated to the empirical data, suggesting its adequacy in capturing the underlying network structure and dynamics.

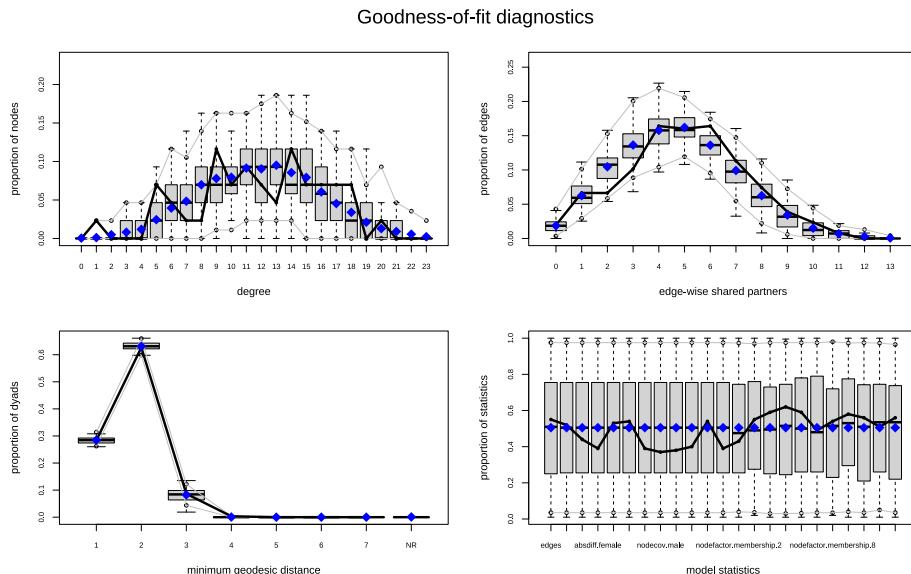


Fig. 3 Goodness-of-fit diagnostics on the final exponential family random graph model

geometrically weighted edgewise shared partners (GWESP) count is incorporated as a higher-order configurational variable. This measure is often implemented to depict intricate network structures and dependency scenarios, and it plays a role in mitigating the issue of model degeneracy by capturing the tendency towards forming closed triangular configurations. However, upon controlling for background variables, the GWESP term's statistical significance diminished and is further attenuated following the addition of categorical factors. This indicates that within the scope of the current study, such structural interconnections are more pronounced within distinct clusters, and socio-physiological factors have a substantive underlying influence.

In the analysis of exogenous effects within the ERGM, it is found that the primary effects associated with gender proportion (nodecov.male and nodecov.female) do not reach statistical significance for either gender. This suggests that the overall presence of each gender within the population under study does not notably affect network formation. Yet, the analysis of absolute attribute differences (absdiff.male and absdiff.female) indicates that increased disparities in female participation rates among sports and physical activities markedly diminish the likelihood of establishing network connections. This phenomenon is mirrored in the participation patterns of males. Consequently, gender stands out as a critical discriminant in the participation patterns of individuals across a range of sports activities. Notably, the data reveals that women's participation in these activities is characteristically more specialized, or exhibits less diversity, in comparison to the participation trends of men.

From the results, years-of-schooling doesn't significantly influence the overall probability of linkage between different sports and physical activities according to the main effect (nodecov.edu). To some extent, this means people in different education levels can play as many sports and physical activities as they want. However, from the result of Model 4 (see the fourth column of Table 3), the coefficient of absdiff term of years-of-schooling (absdiff.edu) is statistically significant, indicating the probability

Table 3 The outputs of ERGM for sports and physical activities and background factors

| Variable | Model | | | | |
|----------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) |
| edges | −0.927*** (0.074) | −2.759*** (0.630) | −2.767*** (0.625) | 0.522 (0.601) | −1.475* (0.693) |
| gwesp.fixed.0.3 | | 1.114** (0.402) | 0.611+ (0.341) | 0.431 (0.325) | 0.325 (0.312) |
| nodefactor.membership.2 | | | −0.417+ (0.245) | | 0.916* (0.453) |
| nodefactor.membership.3 | | | 0.784** (0.256) | | 0.833** (0.274) |
| nodefactor.membership.4 | | | −0.538+ (0.286) | | 0.629 (0.486) |
| nodefactor.membership.5 | | | 0.419 (0.274) | | 0.868** (0.316) |
| nodefactor.membership.6 | | | 0.538+ (0.289) | | 0.930** (0.327) |
| nodefactor.membership.7 | | | 0.912** (0.281) | | 1.080*** (0.322) |
| nodefactor.membership.8 | | | −0.078 (0.349) | | 0.682+ (0.384) |
| nodefactor.membership.9 | | | 0.668* (0.271) | | 0.886** (0.304) |
| nodefactor.membership.10 | | | 0.534+ (0.323) | | 1.015** (0.364) |
| nodefactor.membership.11 | | | 0.540+ (0.289) | | 1.262*** (0.340) |
| nodematch.membership | | | 3.832*** (0.412) | | 3.073*** (0.399) |
| absdiff.male | | | | −1.402** (0.434) | −0.785+ (0.472) |
| absdiff.female | | | | −1.892*** (0.501) | −1.172* (0.546) |
| absdiff.age | | | | −0.439*** (0.118) | −0.513*** (0.139) |
| absdiff.edu | | | | −0.244+ (0.137) | −0.176 (0.149) |
| absdiff.multiple_responses | | | | 1.524* (0.774) | 0.504 (0.820) |
| nodecov.male | | | | −3.958 (8.072) | 3.492 (9.209) |
| nodecov.female | | | | −3.817 (8.051) | 2.970 (9.236) |
| nodecov.age | | | | −0.295*** (0.079) | −0.327** (0.109) |
| nodecov.edu | | | | 0.050 (0.095) | 0.101 (0.127) |
| nodecov.multiple_responses | | | | 8.657 (15.626) | −5.349 (17.873) |
| AIC | 1,078.788 | 1,069.824 | 906.108 | 930.349 | 845.227 |
| BIC | 1,083.594 | 1,079.435 | 968.582 | 988.018 | 955.759 |

+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; Standard errors are in parentheses

of connection decreases for those sports and physical activities with a participation percentage of high education degree and a participation percentage of low education degree accord this coefficient. In other words, the analysis suggests there is a tendency for sports and physical activities to be more popular within certain educational groups, indicating some level of clustering based on education. This reflects the difference of social class in the field of sports and physical activities. After the inclusion of the clustering factor, this item is no longer statistically significant, suggesting that the clustering of sports and physical activities is potentially done with the educational level factor as the main axis.

Observations reveal a pattern with age that parallels the aforementioned educational trends. When controlling for other variables, the absolute difference in age (*absdiff.age*) emerges as statistically significant. Negative coefficients indicate that a greater disparity in age among participants in different sports and physical activities lessens the likelihood of forming connections between these groups. In other words, as the age gap widens among participant groups, the chance of establishing connections across various sports and physical activities diminishes. However, even when the effect of clustering is taken into consideration, the main effect of age (*nodedcov.age*) maintains its statistical significance. This finding suggests that older age cohorts tend to engage in a more limited or less diverse range of sports and physical activities. The implication of this result is that there appears to be a narrower range or fewer options of sports and physical activities that are associated with older age groups. This suggests a potential age-related homogeneity in the choice of physical activities, which could be due to various factors, including but not limited to physical limitations, preferences, or availability of opportunities.

Despite the main effect of the percentage of participation for each sport and physical activity (*nodedcov.multiple_responses*) being initially positive, it becomes negative when clustering factors are taken into account; however, it consistently lacks statistical significance. Correspondingly, while the absolute difference in the participation percentage (*absdiff.multiple_responses*) is initially significant, its significance dissipates upon the control for clustering factors. This suggests that the variability in participant proportions across different sports does not influence the likelihood of connections among those sports when clustering is considered. Consequently, it appears that the clustering factors subsume the distinctive impact of participation proportions on the sports and physical activities, thereby mitigating its influence on network connections.

Furthermore, the examination of the *nodematch* term linked to the sports category (*nodematch.membership*) illustrates a significant homophily effect within individual clusters. The odds of connections between sports and physical activities classified within the same cluster is substantially higher—specifically, 21-fold ($\exp(3.073)$)—when compared to connections between sports and physical activities from disparate categories. Upon investigating the main effect pertaining to sports categories (*node-factor.memberships*), with category 1 serving as the reference, it is evident that nearly all other categories exhibit significantly stronger intra-category associations, and this pattern holds true even after adjusting for other variables. The majority of these categories tend to be specialized and are frequently characterized by exclusivity.

Discuss and conclusion

This study harnesses data from sample surveys on public preferences across a spectrum of 43 sports and physical activities. Through the deployment of network analysis, it classifies these activities and deciphers the intricate topological interrelations between them. It further probes into the multifaceted participation trends in sports and physical activities, examining how these are correlated with the socio-physical characteristics of the individuals involved.

While prior research has often investigated the influence of social networks or social capital on sports engagement primarily at an individual level, this study broadens the scope by employing network analysis to synthesize both macro-level (sports and physical activities) and micro-level (individuals) perspectives, thus considering the interplay between sports and physical activities, and various background factors. The study's outcomes draw attention to the significant role of gender, age, and education in shaping the diversity of sports and physical activities. It reveals a gender gap in participation rates and a tendency for females to participate in a narrower range of activities when compared to males, with certain activities being perceived as gendered. Moreover, the study finds a negative correlation between participant age and the variety of sports activities, indicating a decline in diversity as age increases. Additionally, it identifies a trend among more educated individuals to specialize in sports perceived as prestigious, such as tennis and golf, suggesting that educational attainment influences patterns of sports participation. These findings corroborate and expand upon existing literature on the socio-economic determinants of sports engagement by identifying specific patterns of connectivity and dissociation among various sports and activities, thereby contributing to the discourse on the socio-economic stratification of sports participation, as discussed in the works of Stempel (2005) and Federico et al. (2013).

The study further illuminates a significant clustering effect within sports and physical activities, suggesting that participation choices are not randomly scattered but rather exhibit associative patterns, especially among niche sports. This indicates the existence of sub-communities within the broader sporting community, which may be formed based on shared socio-physical traits.

From a methodological standpoint, the application of network analysis to the study of sports participation provides an intricate understanding of the interconnectivity between different sports and physical activities and how these interrelationships are modulated by the socio-physical attributes of participants. This approach enriches the methodological toolkit of sports studies and is in alignment with the recent methodological advancements in network analysis across various fields, as highlighted by Borgatti et al. (2009), who emphasized its efficacy in unraveling the complex patterns of relationships in data sets.

It is important to note that the analysis level of this study, especially ERGM, is the sports and physical activities. Therefore, the conclusions obtained cannot be directly extrapolated to the individual level. The present study addresses sports and physical activities in China, and some of the results are contextualized, such as the characteristics of the participating population in some emerging sports such as boxing, which differ from the situation in other developed countries. In addition, the expression of some specific options (e.g., "aquatics") in the questionnaire is not reasonable enough and the

understanding of some respondents, especially rural residents, might be skewed, potentially impacting the reliability of the results. Additionally, there are certain unobserved factors that influence participation in sports and exercise programs, which cannot be addressed by the network analysis model itself. This limitation may render the educational attainment variable potentially endogenous, resulting in biased estimations of the coefficient. These issues warrant further refinement in future studies.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s41109-024-00628-0>.

Supplementary material 1.

Supplementary material 2.

Acknowledgements

We would like to thank Ming Yao, Meng Tu, Yuan Shen, Qing Zhang, Yanfei Cao, Hongxu Ji, Guan Jiang, Jianxin He and others. We are also grateful to the three anonymous reviewers and the editors for their valuable comments and suggestions.

Author contributions

The authors jointly designed the project and developed the questionnaire. Xiangyang Bi led the quantitative analysis, Zhanning Sun oversaw the project and qualitative research, and BoRan Hu handled implementation. The initial draft was by Xiangyang Bi, with Zhanning Sun and BoRan Hu contributing to revisions. All authors were engaged in discussing the paper's content.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, Beijing Sport University and/or publication of this article: The data in this article comes from the "Chinese Public Attitudes and Behaviors in Daily Activities Survey" commissioned by the China Basketball Association and conducted by the China Sports Policy Research Center of Beijing Sport University.

Declarations

Conflict of interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Received: 3 January 2024 Accepted: 18 May 2024

Published online: 13 June 2024

References

- Aggio D, Papacosta O, Lennon LT, Ash S, Whincup PH, Goya Wannamethee S, Jefferis BJ (2018) Tracking of sport and exercise types from midlife to old age: a 20-year cohort study of British men. *Eur Rev Aging Phys Act* 15:1–9
- Apostolou M (2015) The evolution of sports: age-cohort effects in sports participation. *Int J Sport Exerc C* 13:359–370
- Baker Y (2017) Methods and applications for mixed graphical models. Doctoral dissertation, Rice University.
- Bandy SJ (2014) Gender and sports studies: an historical perspective. *Mov Sport Sci - Sci Mot* 86:15–27
- Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. *Science* 323(5916):892–895
- Bourdieu P (1978) Sport and social class. *Soc Sci Inf* 17:819–840
- Breznik K, Batagelj V (2012) Retired matches among male professional tennis players. *J Sports Sci Med* 11:270–278
- Chinurum JN, Ogunjimi LO, O'Neill CB (2014) Gender and sports in contemporary society. *J Educ Soc Res* 4(7):25–30
- Clemente FM, Rui SM (2016) Social network analysis applied to team sports analysis. Springer, Switzerland
- Federico B, Falese L, Marandola D, Capelli G (2013) Socioeconomic differences in sport and physical activity among Italian adults. *J Sports Sci* 31:451–458
- Fussell P (1992) Class: A guide through the American status system. Simon and Schuster, New York
- Gantz W, Lewis N (2021) Sports fanship changes across the lifespan. *Commun Sport* 2:1–20
- Gemar A (2021) Social capital networks in sports spectatorship and participation. *Int Rev Soc Sport* 56:514–536
- Gidlow C, Johnston LH, Crone D, Ellis N, James D (2006) A systematic review of the relationship between socio-economic position and physical activity. *Health Educ J* 65:338–367
- Hambrick ME (2013) Using social network analysis in sport communication research. In: Pedersen P (ed) *The Routledge handbook of sport communication*. Routledge, London, pp 279–288
- Hambrick ME (2019) Social network analysis in sport research. Cambridge Scholars Publishing, Newcastle
- Hambrick ME, Marion E (2017) Sport communication research: a social network analysis. *Sport Manag Rev* 20:170–183
- Hardin M, Greer JD (2009) The influence of gender-role socialization, media use and sports participation on perceptions of gender-appropriate sports. *J Sport Behav* 32:207–226

- Harris JK (2013) An introduction to exponential random graph modeling. Sage Publications, Beverly Hills
- Haslbeck J, Waldorp LJ (2018) How well do network models predict observations? On the importance of predictability in network models. *Behav Res Methods* 50:853–861
- Haslbeck JM, Waldorp LJ (2020) mgm: estimating time-varying mixed graphical models in high-dimensional data. *J Stat Softw* 93:1–46
- Hirvensalo M, Lintunen T (2011) Life-course perspective for physical activity and sports participation. *Eur Rev Aging Phys Act* 8(1):13–22
- Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M (2008) ergm: a package to fit, simulate and diagnose exponential-family models for networks. *J Stat Softw* 24:1–29
- Kahma N (2012) Sport and social class: the case of Finland. *Int Rev Soc Sport* 47(1):113–130
- Kovula N (2001) Perceived characteristics of sports categorized as gender-neutral, feminine and masculine. *J Sport Behav* 24:377–394
- Lee JD, Hastie TJ (2015) Learning the structure of mixed graphical models. *J Comput Graph Stat* 24:230–253
- Lusher D, Koskinen J, Robins G (eds) (2013) Exponential random graph models for social networks: theory, methods, and applications. Cambridge University Press, New York
- Nixon HL (1993) Social network analysis of sport: emphasizing social structure in sport sociology. *Soc Sport J* 10:315–321
- Nixon HL (2002) Studying sport from a social network approach. In: Maguire J, Young K (eds) Theory, sport and society. JAI/Elsevier Science Imprint, Amsterdam, pp 267–291
- Norris LA, Didymus FF, Kaiseler M (2020) Understanding social networks and social support resources with sports coaches. *Psychol Sport Exerc* 48:101665
- Onody RN, De Castro PA (2004) Complex network study of Brazilian soccer players. *Phys Rev E Stat Nonlinear Soft Matter Phys* 70:037103
- Paloian A (2012) The female/athlete paradox: managing traditional views of masculinity and femininity. *Applied Psychol OPUS*. https://wp.nyu.edu/steinhardt-appspsych_opus/the-femaleathlete-paradox-managing-traditional-views-of-masculinity-and-femininity
- Ribeiro J, Silva P, Duarte R, Davids K, Garganta J (2017) Team sports performance analysed through the lens of social network theory: implications for research and practice. *Sports Med* 47:1689–1696
- Stempel C (2005) Adult participation sports as cultural capital: a test of Bourdieu's theory of the field of sports. *Int Rev Soc Sport* 40(4):411–432
- Vandervoort AA, Lindsay DM, Lynn SK, Noffal GJ (2012) Golf is a physical activity for a lifetime. *Int J Golf Sci* 1:54–69
- Wäsche H, Dickson G, Woll A, Brandes U (2017) Social network analysis in sport research: an emerging paradigm. *Eur J Sport Soc* 14:138–165
- Yang E, Baker Y, Ravikumar P, Allen G, Liu Z (2014) Mixed graphical models via exponential families. In: Proceedings of the seventeenth international conference on artificial intelligence statistics. PMLR. pp 1042–1050

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.