Sport, how people choose it: A network analysis approach

(Article begins on next page)

20 April 2024

UNIVERSITÀ DEGLI STUDI DI TORINO

Sport, how people choose it. A network analysis approach

Luca Ferreri[*,1,2], Marco Ivaldi[3], Fabio Daolio[4], Mario Giacobini[1,2,5], Alberto Rainoldi[3], and Marco Tomassini[4]

[1]GECO - Computational Epidemiology Group, Department of Veterinary Sciences, University of Torino
[2]ARC[2]S - Applied Research on Computational Complex Systems, Department of Computer Science, University of Torino
[3]Motor Science Research Center - School of Exercise and Sport Sciences S.U.I.S.M., Department of Medical Sciences, University of Torino
[4]Department of Information Systems, Faculty of Business and Economics,University of Lausanne
[5]CSU - Complex Systems Unit, Molecular Biotechnology Center, University of Torino

---
[*]to whom correspondence should be addressed : luca.ferreri@unito.it

Abstract

In order to investigate the behaviour of athletes in choosing sports, we analyse data from part of the We-Sport® database, a vertical social network that links athletes through sports. In particular, we explore connections between people sharing common sports, and the role of age and gender by applying "network science" approaches and methods. Results show a disassortative tendency of athletes in choosing sports, a negative correlation between age and number of chosen sports, and a positive correlation between age of connected athletes. Some interesting patterns of connection between age classes are depicted. In addition, we propose a method to classify sports based on the analyses of the behaviour of people practising them. Thanks to this brand new classifications we highlight links of class of sports and their unexpected features. We emphasise some gender dependency affinity in choosing sport classes.

Keywords: sport classification, network analysis, social network, sport practising, sport aggregation.

Introduction

We-Sport® is a vertical social network that connects athletes and sport professional profiles. It was founded in 2009 as a spin-off at Motor Science Research Center of the School of Exercise and Sport Sciences of University of Turin. We-Sport® allows to organize meetings and training sessions for several sports (there are more that 400 sports registered), to share photos and videos and to find sport facilities. We-Sport® received major investments, both public and private; it won numerous international and national events related to digital innovation and start-ups and it is currently among the top three social networks dedicated to sports in the world. To get updated information related to We-Sport® we refer to the Wikipedia page. The functioning of the platform is straightforward: users enter sports played, places where they practice, their schedule, gender, age and their sporty level. The system suggests those users who best match within 30 km from the place indicated. Currently (January 2014) about 50,000 people around the world are We-Sport® users. Moreover, the social network has approximately 23,000 page views and 5,000 unique visitors per month. Since its creation We-Sport® had 200,000 unique visitors for nearly 1 million page views. According to Alexa (Alexa Internet company provides traffic data, global rankings and other information on 30 million websites; Alexa ranking is one of the indicators used for monitoring of traffic data and global web rankings) presently We-Sport®

1

is among the top 8000 sites in Italy and among the first 300,000 in the world as number of visits (Alexa Internet, 2014). We-Sport® visitors come from every country in the world, the 10 countries which have the highest number of visitors on We-Sport® are, in order: Italy, Hong Kong, United States, France, Germany, Spain, Turkey, United Kingdom, Brazil, Switzerland. Regarding a demographic overview, by using Google Analytics we observe that the percentage of users between 18 and 24 years old is 18%, 32% between 25 and 34 years old, 25% between 35 and 44 years old, 14% between 45 and 54 years old, 7% between 55 and 64 years old and 4% are people with more than 65 years old. 53% are male and 47% are female. For more information on the use of Google Analytics we refer to works of Plaza (2009, 2011); Amin Omidvar, Mirabi, and Shokry (2011) and of Clifton (2010).

Nowadays we observe a flurry of research on social networks, the field is a relatively old one, having been created in the 1930s in the US by researchers coming from cognitive and social psychology. They supposed that a relation between two people can structurally be viewed as an edge connecting two points in a space, each representing a person. The set of relationships thus obtained defines a sociogram, which is a graph that today is called a social network. In the following decades many quantitative ways of analysing such structures were created, giving rise to the field of sociometric analysis. Many of these ideas and methods are still used today. The main developments until the sixties are well described in a book by Scott (2013) to which the reader is referred for details. Other useful references and details can be found in the main book in the field, (Wasserman & Faust, 1994).

In the last two decades the introduction of new communication technologies has notably enlarged the concept of social network. While the first sociograms were necessarily small and had to be built by hand through patient and error-prone survey work, today an enormous amount of data is available online and can be scanned quickly using suitable software. Internet-mediated social networks such as Facebook and Twitter have hundreds of millions of participants and a global structure that is almost impossible to compute and describe as a whole. This fact, has brought many new actors into the field, which is today one of the most active and fast-developing, since the network concept permeates almost all disciplines and is certainly not limited to social interaction. Recent books describing this field are those of Newman (2010), and of Barrat, Barthélemy, and Vespignani (2008).

In this paper, classical approaches in social network analysis were adopted to investigate the aggregative behaviour of sportsmen, ranging from how athletes choose different sports and the potential connections between people sharing common sports, to the role played by

age and gender. Moreover, applications specifically designed for the interpretation of this peculiar social structure were proposed. Furthermore, given some sports classifications built up by using different criteria, such as seasonal or physiological classification, we investigate how these sets of disciplines are bound together.

In this context, the analysis of virtual communities is crucial to fully understand social and behavioural mechanisms related to the aggregation. Existing literature (Ridings & Gefen, 2004) theorises that people join virtual communities to exchange information and/or social support. Theories of broader Internet use have indicated both entertainment and friendship seeking as motivational forces. The reasons for people to choose a particular sport rather than another have been investigated extensively in the past (Bourdieu, 1978) albeit without using data from virtual communities. In particular, many studies have been performed in relation to children aggregative sports behaviour (Whitehead, 1993), trying to understand why children start a sport and why at some point of their lives they decide to stop practising it. Dishman, Sallis, and Orenstein (1985) recommended specific study areas that could help in understanding what motivates people to become physically active and help in developing ways to increase activity. In fact, they categorised determinants related to the adoption and maintenance of physical activities: personal factors, environmental factors, or characteristics of the exercise. In some works the analysis of sports choice was done by evaluating the perception of dangerousness (Russell, 2005). Some other analysis, (Ko, Claussen, & Park, 2008), were conducted to reveal the commercial factor in choosing sports. This study includes both the cost of the practice (e.g. specific equipment, travels, teachers) and the affiliation cost to groups of players (e.g. tennis, golf, sailing club). Literature is also rich in works depicting trends associated with specific social groups (by age groups or specific behaviours). Several studies have investigated the differences in sport related to gender, we cite the work of Eccles and Harold (1991) as the most complete. Eccles and collaborators conclude that gender differences in children's attitudes toward sports are quite strong and emerge at a very young age. Furthermore, they suggest that gender differences are more related to the gender-role in socialisation than to "natural" attitudinal differences. Our social network analysis could confirm or improve their conclusions while other studies, also related to the sport habits as consequence of gender behaviour, have been conducted by Gard (2003) and Lantz and Schroeder (1999).

With respect to the methods of classification of sports, there have been several proposals as this issue is one of the most controversial issue related to the sport world. Over the years, many authors de-

scribed different possibilities but all refer to a theoretical and a priori method. A number of methods were proposed by analysing either the technical aspect only, or the physiological aspect (Bellotti et al., 1978; Meinel, 1984; Berger & Hauptmann, 1985; Starosta, 1987a, 1987b; Merni, 1988; Dal Monte & Faina, 1999). Csizma, Wittig, and Schurr (1988) proposed a classification related to the distinguishable gender association, Chelladurai (1992) in relation to managerial implications, Chen and Zhang (2005) and Lu, Yang, and Luo (2009) on the basis of technical and educational parameters. However, most classifications are based on the fundamental objectives of the technical basis or on the physiology of the exercise (Scotton, 2004; Mitchell, Haskell, & Raven, 1994; Riemer & Visio, 2003; Mitchell et al., 1985; Committee on Sports Medicine, 1988; Ainsworth et al., 1993, 2010, 2011). More recently, some works have proposed a classification on the basis of information such as touristic or commercial (see (Gammon & Robinson, 2003) or beginning to use big data information from the web through technological device (Parkka et al., 2006). Our work could provide a classification based on the actual choice of practiced sports by people, a choice that could be related with multiple factors (age, sex, wealth, habits, education, geographic origin). Instead of defining an *a priori* approach classification of sports on theoretical grounds, we propose a *bottom up* approach based on the analyses of the behaviour of people practising sports. This approach attempts to classify sports according to the response to the eight fundamental questions of the Rhetoric (called circumstances or *loci argumentorum* (Tommaso d'Aquino, Summa Theologiae, 1265-1274), namely: *quis*: who? - with whom; *quid*: what? *quando*: when? *ubi*: where? *cur*: why? *quantum*: how much? *quomodo*: how? *quibus auxiliis*: with what kind of equipment?

For questions "who", "with whom", we consider if a sport is played mostly in solitude, in pair or in team, considering that sports that are practised in groups can also be practised in solitude, especially with regard to non-competitive practice, and, conversely, sports that can be done in solitude can also be enjoyed in company.

The questions "Where?" and "When?" resulted in a division of sports in three distinct sub-questions. The first is related to the environment, seen either as an artificial sport facility (e.g. an hockey field), or a natural-but-human-modified environment (e.g. a golf pitch), or a wild or pristine one (considering that if a sport is practised outdoors without special facilities, it does not mean that it can not be played in specific structures). The second one is related to the type of field, e.g. air, water, land, etc. Finally, the classification of a sport was linked to the season during which the sport is usually practised, namely hot

or cold season.

The question "Why?" was translated to the question "What is the goal?", whereas the question "How" relates to the neurophysiological aspects required for the athlete's performance. Then, the question: "With which kind of equipment?" deals with the necessary equipment to play the sport.

The characteristics of the sports in the classification are prevalent and not exclusive. What we are herein proposing is a classification according to objective features; nevertheless we are aware that every inclusion of a sport is questionable and worthy of further investigation. What we propose is a first milestone that will require further validation, improvements and additional specifications. The proposed classifications are summarised in Table 1.

This study was carried out with the consensus of participants, explicitly provided when accepting the terms of use of the We-Sport® social network. The study received institutional ethics approval.

The Network

A network, or a graph, is defined as a set of nodes and a set of links, or edges, connecting pairs of nodes (see Figure 1 a)). We-Sport® social network is an example of graph in which the users, teams, sportive disciplines, and locations are the nodes and their interactions are the edges. We-Sport® naturally defines a tetrapartite graph – a graph in which nodes are partitioned in four sets and connections are only between pairs of nodes belonging to different sets (see Figure 1 b)) – having users belonging to different teams and playing different disciplines in different places. Indeed, the connection between two sports or athletes is meaningless for the We-Sport's® original aim: connecting athletes through their favourite sports. Even if the tetrapartite representation of the We-Sport® social network , see Figure 1 b), is the most accurate, in this article we consider only nodes belonging to two sets: athletes and sports. In other words, we consider the bipartite (also known as two-mode, (Wasserman & Faust, 1994)) version of the network, and we study its structure at December 2010. In particular, the network accounts for 1678 users, 238 sportive disciplines and the 6107 connections between them. For a graphical representation of the reduced bipartite network of We-Sport® we refer to Figure 1 c).

A classical way to quantify the role of a node (i.e. an athlete or a sport) in a graph is to consider its number of connections, i.e. the degree of the node. Managing and visualising this information in networks having hundreds of nodes can be really hard. However, some statistics could help us in revealing the underlying structure (Barrat

Item of classification	Name	# sports	Description
Number of partner	solo	146	Sports playes in solitude
	1 vs 1	34	Sports played in couple
	team	58	Sports played in team
Place	close	95	Sports played in facilities
	open MOD	62	Sports played outdoor but in artificial situation
	open NAT	81	Sports played outdoor in natural situation
Environment	air	11	Sports played in air environment
	water	20	Sports played in water
	ground	170	Sports played on ground
	ice-snow	24	Sports played on snow or ice
	sea	12	Sports played on beach, sand or sea
	composite	1	Sports composite
Season	cold	17	Sports played during cold season
	hot	22	Sports played during hot season
	indifferent	200	Sports played during hot and cold season
Instrument	specific equipment	52	Sports played with specific equipment
	locomotion mean	73	Sports played with a mean of locomotion
	balls	53	Sports played with balls or similar
	nothing	60	Sports played without an instrument
Target	time	61	Sports played with time as goal
	points	72	Sports played with points as goal
	individual target	39	Sports played with individual goal
	aesthetics	46	Sports with an aesthetic judgment
	opponent defeat	12	Sports with physical defeat of the adversary as goal
	measure	7	Sports played with misure as goal
Neuro-physiological aspects	endurance	59	Sports based on musculoskeletal and cardio resistance
	strenght	9	Sports based on muscular strength
	speed	5	Sports based on speed
	precision	21	Sports based on accuracy
	acrobatics	34	Sports based on acrobatics
	rapidity	52	Sports based on rapidity and reaction time
	others	58	Sports based on other aspects or mixed aspects
Other aspects	martial art	14	Martial arts and fighting
	dance	7	Dances
	elitist	10	Elitist and expensive sports
	brain	5	Intellectual sports
	gymnastic	3	Gymnastics
	track and field	7	Track and field sports
	composite	2	Composite sports
	animals	4	Sports with animals
	ambient gym	10	Gym and related
	engine	8	Sports with vehicles
	holistic gym	3	Holistic gymnastics

Table 1: The adopted classifications, the class names, the number of sports in each class, and a brief description of the class.
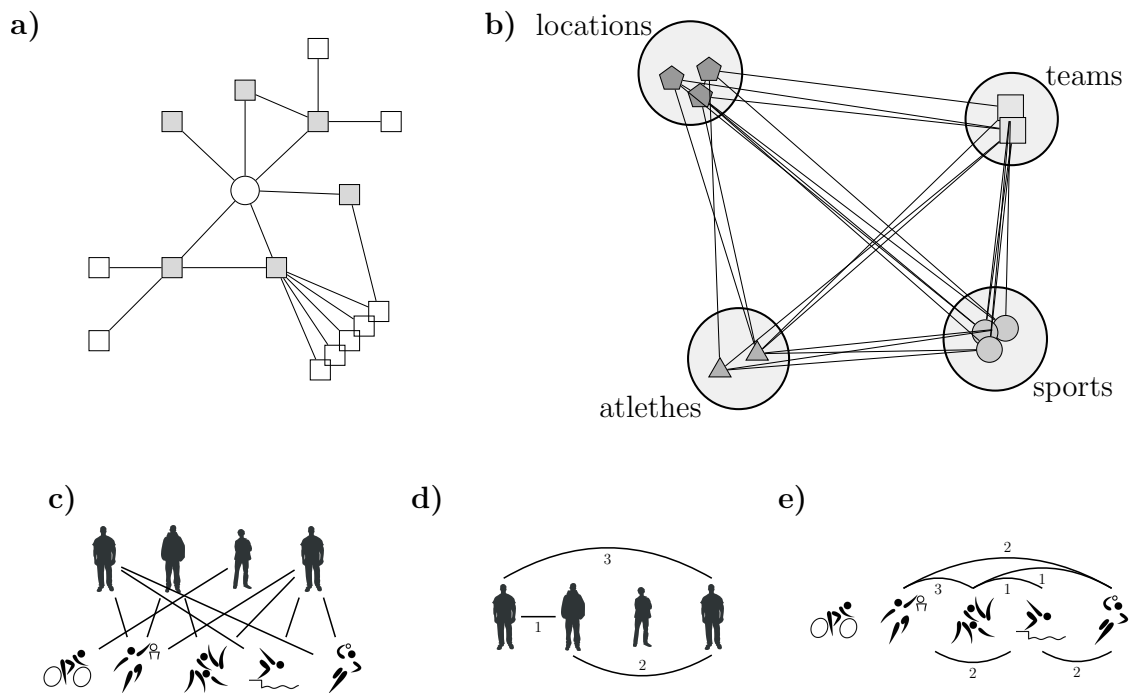
Figure 1: a), an example of graph. The nodes belonging to the neighbour-hood of the node circle are depicted in light grey. The circle node has a degree equal to six. b), A tetrapartite graph. Nodes are divided in four groups. Edge between nodes in the same group are not allowed. c) the bipartite structure of the social network: no edges between nodes of the same family. On the top the athletes' class, on the bottom the sports' class. The weighted projection over athletes, d), and over sports e).

et al., 2008; Scott, 2013; Newman, 2010; Wasserman & Faust, 1994).

Results

Network Characterisation

Keeping in mind that the number of sports eligible by a user was arbitrarily limited to 33 by the administrator of the system (in order to avoid users selecting more sports than those they really practice), we observe that half of the considered athletes choose no more than 3 sports (i.e. median value). It also appears that half of the 238 sports are chosen by no more than 5 users while Jogging was the most chosen sport (more than 30% of users), followed by Five-a-side Football, Swimming, Tennis (25% of users), etc. However, a typical and informative way to explore the overall node features is to consider the degree distribution $p(k)$ of the entire network, which, in our case, provides the frequency of an athlete (a sport) connected with k sports (athletes). In particular, the fitting of the empirical distribution with a theoretical function could help in evaluating nodes behaviour and gives important clues necessary for the understanding of the nodes aggregation process. In our data, the degree distribution of sport-nodes and athlete-nodes is heavy tailed, thus defining an heterogeneous graph (Barrat et al., 2008). Following the suggestions of Clauset, Shalizi, and Newman (2009), we fit the empirical data to several functional forms: Power-Law (PL), Weibull (WEI), Exponential (EXP), Log-Normal (LN), Yule (YU), Poisson (POI), and Negative-Binomial (NB). In fact, we are interested in correctly describing the tail of the distribution – the most influencing part of the distribution since it drives the variability of data and therefore highly affects the process occurring on graph (for more details on this wide topic we refer to (Barrat et al., 2008; Newman, 2010)). Results show that the estimated scaling parameters of the PL for sports (athletes) nodes having a number of connections larger than 15 (6) is 1.9 (3.48) with a acceptable goodness-of-fit ($p > 0.05$) (Clauset et al., 2009). Moreover, performing a likelihood-ratio-test later interpreted as suggested by Vuong (1989) we find evidences that the PL has to be preferred ($p < 0.05$) to the BN, POI, EXP in fitting the degree distribution of sport-node. Moreover, the PL fit of the degree distribution of athletes-nodes has to be ($p < 0.05$) preferred to BN and POI while the WEI, LN, and YU fit should be ($p < 0.05$) preferred to the PL fit. The fact that, for athletes-nodes, the WEI, LN, and YU are preferred to PL should not surprise. The arbitrary maximum number of sports that an athlete could choose bounded to 33 inhibit the formation of a

real heavy tail thus making distribution belonging to the exponential family more suitable in fitting the data.

Users Behaviour

We try to detect possible behavioural pattern in our data. A first question was to explore if gender is a factor when athletes choose sports. In this context, a non parametric test – Mann-Whitney test (Zar, 2010) – allow us to conclude that gender is not a factor influencing the number of sports chosen by users.

We also investigated if the number of sports chosen by an athlete is related to the popularity of the sports (quantified by the number of users choosing the sport). Let's consider the set of athletes playing k disciplines: we want to investigate the correlation between k and the mean number of people playing the k considered sports. Since data show evidence of deviation from normal distributions, to quantify the correlation we use a non parametric method – the Spearman rank correlation (Zar, 2010). Results ($r = -0.9560$ with $p < 0.01$) point out a negative correlation, see Figure 2 a). We further explore the athletes behaviour by considering the correlation between the number of sports that connected people choose. Let us recall that in We-Sport$^{\textregistered}$ network athletes are not allowed to directly connect through friendship relationships. However, it is possible to detect the potential set of people that an individual could encounter thanks to its connections to sports. This can be studied by generating the weighted projection of the bipartite graph, (Zhou, Ren, Medo, & Zhang, 2007), over the athlete-nodes. This graph is formed by athlete-nodes only and a pair of these nodes are connected if before the projection they shared at least a common sport. Moreover, this graph is said weighted because we assign to each edge a numerical attribute indicating the number of sports shared by the couple of sportsmen, we refer to Figure 1 d) for an example. In this projected graph we find an assortative behaviour, i.e. people tending to connect with similar ($r = 0.8701$ with $p < 0.01$), see Figure 2 b).

Age Pattern

Keeping in mind that only users older than 18 could subscribe to the We-Sport$^{\textregistered}$ network, we explore the role of age in choosing sports. We find evidence of a negative correlation ($r = -0.6442$ with $p < 0.001$) between the age of athletes and the number of sports that they choose, see Figure 2 c). Finally, working on the weighted-projected graph, we observe a positive association ($r = 0.8304$ with $p < 0.001$) between
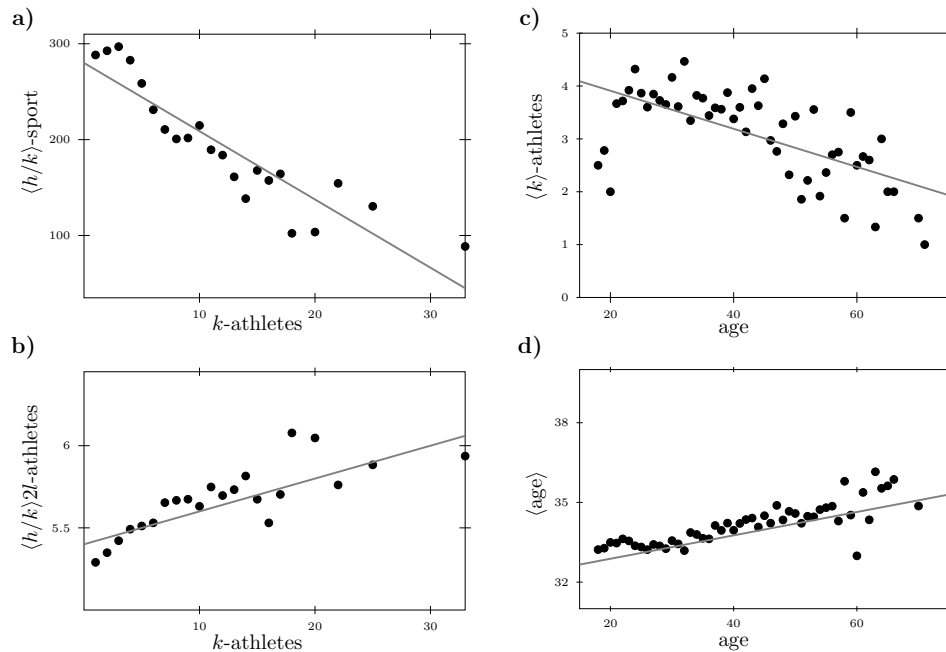
Figure 2: a) On the horizontal axes the number of sports chosen by users while on the vertical axes the mean number of players that play such sports. A clear negative correlation is observed. b) On the horizontal axes the number of disciplines played by users and the number of sports played by others with whom they are connect among sports. An assortative manner could be observed. c)The age of sportsmen and their averaged number of sports. The averaged number of sports seems to decline as the age of athletes increase. d) The age of sportsmen and the averaged age of those connected with them. A linear, positive, correlation could be depicted.

the age of tied athletes, Figure 2 d). We further explore the age association pattern by quantifying if pairs of age groups are more connected than expected by the graph topology. Mathematically this could be evaluated by

$$E(a_i, a_j) - \frac{s_i \cdot s_j}{\sum_i s_i} \tag{1}$$

where $E(a_i, a_j)$ is the number of connections between nodes belonging to age group a_i and a_j, taking into account the weight attribute of edges, $s_i = \sum_j E(a_i, a_j)$ is the number of connections connecting pairs of nodes belonging to age group a_i and a_j, and $\frac{s_i \cdot s_j}{\sum_i s_i}$ is the expected number of connections between the pair of age group. This difference indicates whether the number of edges is just a consequence of the graph topology or if it is independent from it, allowing us to reveal patterns between age classes. Finally, it is worth to underline that Equation 1 is not far from what Newman and Girvan (2004) use to define the modularity function, an important measure to detect communities in large networks. Results of Equation 1 are in Figure 3 a).

Communities, Sport Classifications and Mixing Patterns

On a graph a community can be defined as a subset of nodes having more connections among them with respect to those between them and the other nodes in the network (Newman, 2010). In the We-Sport® network community detection could be of great interest to recognise groups of sports, or of athletes, sharing some topological features and/or some behavioural patterns. Several algorithms have been developed in order to automatically identify communities in networks (Fortunato, 2010). Unfortunately, such algorithms, when applied to our network, in its original or projected versions, give too poor results to draw any significant conclusion. For instance, communities detection on the weighted projection over sport (i.e. the reduced graph in which pair of sports are connected with an edge of weight w if w athletes choose both sports) provided results comparable to those of random graphs (Guimerà, Sales-Pardo, & Amaral, 2004). Our main explanation of this fact is that our dataset is too small to have a meaningful signal. However, we refer to a future work for a deeper analysis. Here, working on the weighted projection on sports of the bipartite graph, we investigate if group of sports are more connected than expected by the graph topology. To explore this question, again, we use Equation 1.

From the application of Equation 1 on our sport classification we can highlight a weak association, weaker than expected by the graph topology, between sports played in composite environment and those played in teams. We also observe a similar behaviour between sports played in composite environment and sports played with balls and with physiological features related to the rapidity. Conversely, there is a stronger than expected relation between sports played alone and sports played outdoors and with physiological features related to the endurance.

We further apply Equation 1 to reveal unexpected association of sport groups with the gender of athletes. Results are in Figure 3 b).

Discussion and Conclusions

The number of sports chosen by users is heterogeneously distributed and it is not gender related: we observe a huge number of users selecting only few sports and a smaller number of users that select a large number of sports. A similar behaviour is also observed for the number of players that each sport owns. We expected more uniform and symmetric behaviours in choosing sports: we expected to find people behaving more similarly, with a large number of athletes selecting a number of sports near to the mean and sports that have a number of practitioners homogeneously and symmetrically distributed around the mean.

The negative association observed between athletes and sports popularity (cfr. Figure 2 a) suggests that athletes who practice few sports are associated with sports practised by many athletes. Moreover, athletes playing a large number of sports, on average, are associated with less popular sports. This feature is also the cause of the athletes assortativity: the larger the number of sports I play, the larger the mean number of sports played by users connected with me through these sports. An explanation for this feature, observed in many other social networks, could be that very often users have few time or resources to play sports. Hence, they focus on those sports that are the easiest to be played: with high probability, those that are the most common, that avoid the lack of partner problem, that are the most feasible, thanks to the presence of facilities over the area, and that sportsmen are used to play, because of cultural or educational biases. People that have more time, and resources, behave in a more complex way selecting both common and rarer sports. These people select common sports as first choice but then improve their set of choices in a more targeted way. Policy makers, public health employees, and sport scientists should be aware of that behaviour in
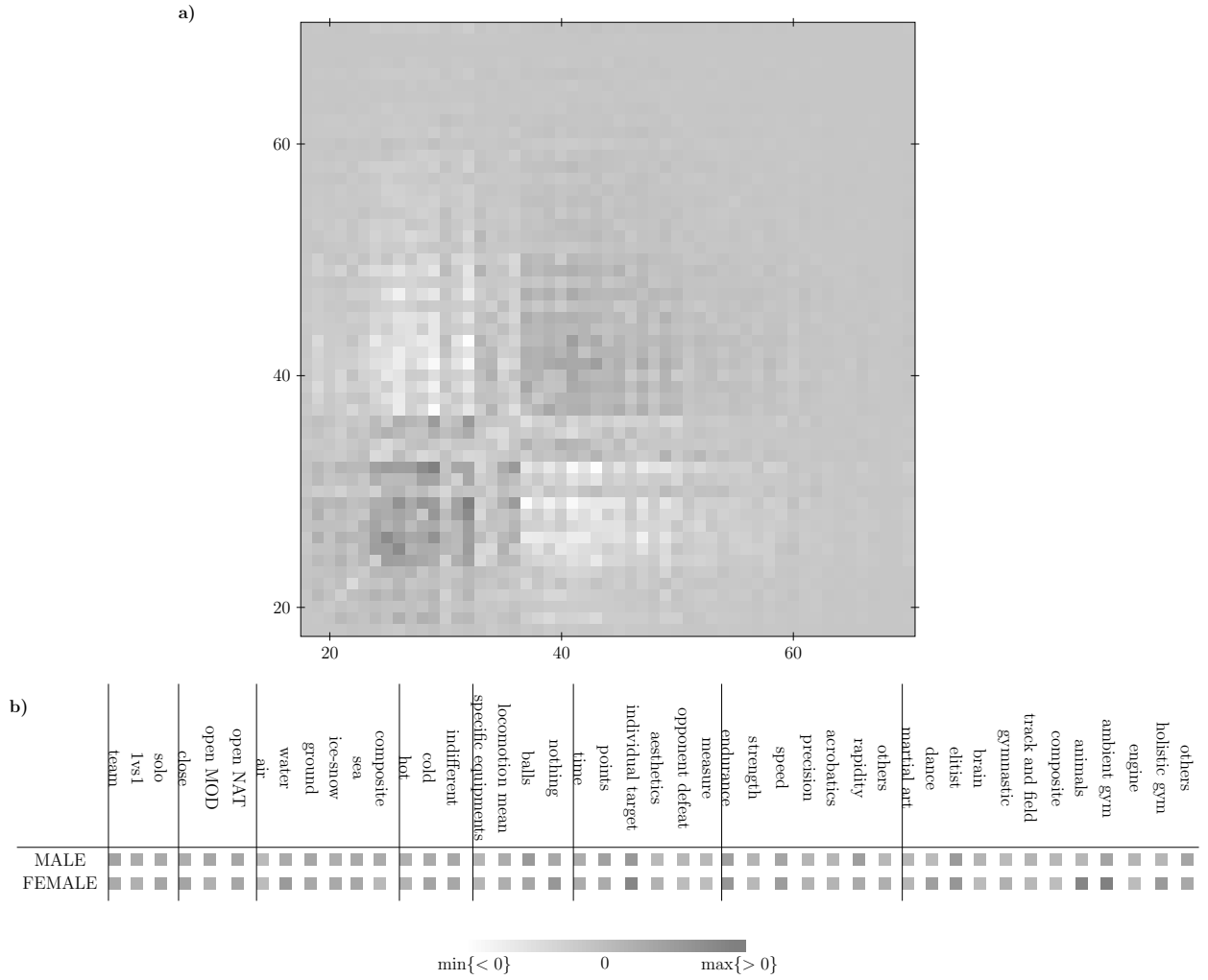
Figure 3: a) The spread between the number of connections within age classes with respect to the expected value. Age group patterns could be distinguished. b) The spread between the number of connections within gender and group of sports with respect to the expected value. Some gender behaviour could be identified.

order to improve their policies, their choices and their work in the support of the diffusion of sports. These findings allow us to speculate that sports with low diffusion have intrinsic characteristics that make them attractive for small groups of people just due to their very low diffusion. The vicious circle is then fed by the fact that a sport with a low diffusion is chosen by the few players appreciating the exclusivity "plus" in such activities. Once that sport reaches high diffusion, it is selected by people liking sports with high diffusion, while the group that chooses the sport with low diffusion will tend to replace it with another sport with characteristics similar to the one just lost.

The negative correlation between the age of sportsmen and the number of sport chosen found in our dataset is in line with previous findings (Barber & Havitz, 2001; Breuer & Wicker, 2008; Downward, 2007; Downward & Riordan, 2007; Bauman, Sallis, Dzewaltowski, & Owen, 2002; García, Lera-López, & Suárez, 2011; Farrell & Shields, 2002; Humphreys & Ruseski, 2006; Moens & Scheerder, 2004; Scheerder, Vanreusel, & Taks, 2005; Breuer, Hallmann, Wicker, & Feiler, 2010; Wicker, Breuer, & Pawlowski, 2009) and can be attributed to several causes (Stamm & Lamprecht, 2005). We also demonstrated a positive correlation between the age of athletes and the age of those connected with them by sports, (cfr. Figure 2 d)). We think that this result is a consequence of the cohort effect affecting choosing behaviour, (Stamm & Lamprecht, 2005). This result is also confirmed by Figure 3 a). We-Sport$^{\circledR}$ users aged between 20 and 40 are more connected, through sports, between them than what we expected by a random graph with the same topology. Moreover they are less connected with users aged between 40 and 60 than expected. This result reveals that sports could be roughly divided in two groups: one among which young user are connected to each other, and another which connect older sportsmen. To our knowledge, this finding has not yet been evaluated in relation with sport habits. However, in other contexts, similar matrices have been estimated revealing pattern not far from our estimates, we refer for instance to (Mossong et al., 2008).

Results on the mixing patterns of groups of sports show some interesting features. For instance, we observe a strong connection among sports played in solitude (like running or cycling), and another strong connection between solo sports and sports played on ground (like mountain biking, golf or hiking). Another strong connection is between sports played on ground and sports played without specific seasonal characteristics (like snowboarding for cold season or beach volley for hot season). Extending such an analysis on a larger dataset will allow us to know, with a high degree of reliability, which kind of sports a person practices or would like to practice just knowing the

sport group already attended from the person.

The most interesting result comes from Figure 3 b), that shows the different preferences for gender. As it can be seen, males prefer sports played in team (like volley) and performed outdoors (like golf), played with balls (like soccer), with score or individual goals as target (like archery), and that have physiological aspects related to the muscular strength, speed, and quickness (like climbing or karate). They also prefer sport related to gymnastic environment (like body building). On the contrary, females show preferences for sports played alone (like jogging), indoors (like fitness), in water (like swimming), without specific tools (such as running), and in many cases related to the hot season, with a strong propensity for sports with individual goals linked to the endurance (like running or nordic walking). Females also have a significant propensity to play sports in the dance and in the gymnastics field, in particular in relation to holistic gymnastic (such as yoga). They also show a preference for sports played with animals (like horsemanship). In general, the marked difference between genders is found in the female little tendency for sports played in team or played in pairs (like basketball or tennis), for composite sports (like triathlon) and for sports played with balls or similar and that have a score (target, measure, or opponent defeat like shooting, power lifting or martial arts). They also have little affinity with sports related to the speed (like short track or skiing) and with engines (karting) and for intellectual sport (like wargame). Conversely, male show more cross-cutting, and in general they do not like sports related to the dances, performed with the animals and with holistic world. They like sports in water environment (swimming or aquagym) less than women.

This paper is, to our knowledge, the first work where the "Network Science" methods have been applied to the analysis of the behaviour of athletes in choosing sports. Our hope is to have made a contribution in this field both presenting new methods and new results. However, a number of challenging questions are still open. Therefore, in the future we would like to extend our work in several directions: to study the tetrapartite structure of the network as a whole, to deeper analyse the community in the graph, to study the evolution of the network in time, and to understand how the geographical variables influence its features.

References

Ainsworth, B., Haskell, W., Herrmann, S., Meckes, N., Bas-

sett Jr., D., Tudor-Locke, C., ... Leon, A. (2011). 2011 compendium of physical activities: A second update of codes and met values. *Medicine and Science in Sports and Exercise*, *43*(8), 1575-1581. doi: 10.1249/MSS.0b013e31821ece12

Ainsworth, B., Haskell, W., Leon, A., Jacobs Jr., D., Montoye, H., Sallis, J., & Paffenbarger Jr., R. (1993). Compendium of physical activities: Classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise*, *25*(1), 71-80. doi: 10.1249/00005768-199301000 -00011

Ainsworth, B., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., ... Leon, A. S. (2010). Compendium of physical activities: an update of activity codes and met intensities. *Medicine & Science in Sports & Exercise*, *32*. Retrieved from `http://journals.lww.com/acsm-msse/Fulltext/2000/09001/Compendium_of _Physical_Activities__an_update_of.9.aspx`

Alexa Internet, I. (2014). *Alexa top 500 global sites.* Retrieved from `http://www.alexa.com/siteinfo/we-sport.com`

Amin Omidvar, M., Mirabi, V. R., & Shokry, N. (2011). Analyzing the Impact of Visitors on Page Views with Google Analytics. *International Journal of Web & Semantic Technology*. Retrieved from `http://airccse.org/journal/ijwest/papers/0111ijwest02.pdf`

Barber, N., & Havitz, M. (2001). Canadian participation rates in ten sport and fitness activities. *Journal of Sport Management*, *15*(1), 51-76. Retrieved from `http://www.scopus.com/inward/record.url?eid=2-s2.0-0035529099&partnerID= 40&md5=7fba512d9874985049506c6cfda54937`

Barrat, A., Barthélemy, M., & Vespignani, A. (2008). *Dynamical processes on complex networks.* Cambridge, UK: Cambridge University Press.

Bauman, A., Sallis, J., Dzewaltowski, D., & Owen, N. (2002). Toward a better understanding of the influences on physical activity: The role of determinants, correlates, causal variables, mediators, moderators, and confounders. *American Journal of Preventive Medicine*, *23*(2 SUPPL. 1), 5-14. doi: 10.1016/S0749-3797(02)00469-5

Bellotti, P., Benzi, G., Dal Monte, A., Donati, A., Matteucci, E.,

16

& Vittori, C. (1978). Classificazione degli sport e determinazione dei mezzi di allenamento [Classification of sports and determination of training means]. *Atleticastudi*, *3*, 29-46. Retrieved from `centrostudi.fidal.it/index.php ?option=com_sobi2&sobi2Task=sobi2Details&catid= 415&sobi2Id=602&Itemid=3`

Berger, J., & Hauptmann, M. (1985). Classificazione degli esercizi fisici [Physical exercises classification, original in germany]. *Scuola dello Sport*, *IV*(2), 36-41.

Bourdieu, P. (1978). Sport and social class. *Social Science Information*, *17*(6), 819-840. doi: 10.1177/ 053901847801700603

Breuer, C., Hallmann, K., Wicker, P., & Feiler, S. (2010). Socio-economic patterns of sport demand and ageing. *European Review of Aging and Physical Activity*, *7*(2), 61-70. doi: 10.1007/s11556-010-0066-5

Breuer, C., & Wicker, P. (2008). Demographic and economic factors influencing inclusion in the german sports system: A microanalysis of the years 1985 to 2005. *European Journal for Sport and Society*, *5*(1), 35-43.

Chelladurai, P. (1992). A classification of sport and physical activity services: implications for sport management. *Journal of Sport Management*, *6*(1), 38–51. Retrieved from `http:// www.cabdirect.org/abstracts/19921899734.html`

Chen, X.-l., & Zhang, B.-q. (2005). Thinking of the classification of sport. *Journal of Xi'an Institute of Physical Education*, *6*. Retrieved from `http://caod.oriprobe.com/articles/9371811/ Thinking_of_the_Classification_of_Sport.htm`

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661-703. doi: 10.1137/070710111

Clifton, B. (2010). *Advanced web metrics with Google Analytics.* Sybex.

Committee on Sports Medicine. (1988). Recommendations for participation in competitive sports. *Pediatrics*, *81*(5), 737-739. Retrieved from `http://pediatrics .aappublications.org/content/81/5/737.abstract`

Csizma, K. A., Wittig, A. F., & Schurr, K. T. (1988). Sport stereotypes and gender. *Journal of Sport & Exercise Psychology*, *10*(1), 62 - 74. Retrieved from

http://search.ebscohost.com/login.aspx?direct=
true&db=psyh&AN=1989-13980-001&site=ehost-live

Dal Monte, A., & Faina, M. (1999). *Valutazione dell atleta [Athlete evaluation, original in italian].* Torino, Italy: UTET.

Dishman, R., Sallis, J., & Orenstein, D. (1985). The determinants of physical activity and exercise. *Public Health Reports*, *100*(2), 158-171. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0021828473&partnerID=40&md5=258e7502563bd847f6c8a1e09064b400

Downward, P. (2007). Exploring the economic choice to participate in sport: Results from the 2002 general household survey. *International Review of Applied Economics*, *21*(5), 633-653. doi: 10.1080/02692170701474710

Downward, P., & Riordan, J. (2007). Social interactions and the demand for sport: An economic analysis. *Contemporary Economic Policy*, *25*(4), 518-537. doi: 10.1111/j.1465-7287.2007.00071.x

Eccles, J. S., & Harold, R. D. (1991). Gender differences in sport involvement: Applying the eccles' expectancy-value model. *Journal of Applied Sport Psychology*, *3*(1), 7-35. doi: 10.1080/10413209108406432

Farrell, L., & Shields, M. (2002). Investigating the economic and demographic determinants of sporting participation in england. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *165*(2), 335-348. doi: 10.1111/1467-985X.00626

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3), 75–174. doi: 10.1016/j.physrep.2009.11.002

Gammon, S., & Robinson, T. (2003). Sport and tourism: A conceptual framework. *Journal of Sport & Tourism*, *8*(1), 21-26. doi: 10.1080/14775080306236

García, J., Lera-López, F., & Suárez, M. (2011). Estimation of a structural model of the determinants of the time spent on physical activity and sport: Evidence for spain. *Journal of Sports Economics*, *12*(5), 515-537. doi: 10.1177/1527002510387080

Gard, M. (2003). Moving and belonging: Dance, sport and sexuality. *Sex Education*, *3*(2), 105-118. doi: 10.1080/14681810309037

Guimerà, R., Sales-Pardo, M., & Amaral, L. A. N. (2004, Aug). Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E*, *70*, 025101. doi: 10.1103/PhysRevE.70.025101

Humphreys, B. R., & Ruseski, J. E. (2006). Economic determinants of participation in physical activity and sport. *International Association of Sports Economists - Unpublished Working Paper Series*(06-13).

Ko, Y., Claussen, C., & Park, H. (2008). Action sports participation: consumer motivation. *International Journal of Sports Marketing and Sponsorship*, *9*(2), 111-124. Retrieved from `http://www.scopus.com/inward/record.url?eid=2-s2.0-70349216843&partnerID=40&md5=aa78effb16e3df56c014178457e2435e`

Lantz, C. D., & Schroeder, P. J. (1999). Endorsement of masculine and feminine gender roles: Differences between participation in and identification with the athletic role. *Journal of Sport Behavior*, *22*(4), 545 - 557. Retrieved from `http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1999-15471-004&site=ehost-live`

Lu, C.-f., Yang, W.-x., & Luo, X.-b. (2009). Analysis of the classification of sport disciplines and suggestions for adjustment. *Journal of Physical Education*, *4*. Retrieved from `http://en.cnki.com.cn/Article_en/CJFDTOTAL-TYXK200904003.htm`

Meinel, K. (1984). *Theories of the movement [Original in germany]*. Sports Society Press.

Merni, F. (1988). La valutazione delle tecniche sportive [Sport techniques evaluations, original in italian]. *Scuola dello Sport*, *VII*(13), 23-37.

Mitchell, J. H., Blomqvist, C. G., Haskell, W. L., James, F. W., Miller, H. S., Miller, W. W., & Strong, W. B. (1985). Classification of sports. *Journal of the American College of Cardiology*, *6*(6), 1198–1199. doi: 10.1016/S0735-1097(85)80200-X

Mitchell, J. H., Haskell, W. L., & Raven, P. B. (1994). Classification of sports. *Journal of the American College of Cardiology*, *24*(4), 864–846. Retrieved from `https://www.scopus.com/inward/record.url?eid=2-s2.0-0027993941&partnerID=40&md5=c73e14f1786b0951ef062567d55aed05`

Moens, M., & Scheerder, J. (2004). Social determinants of sports participation revisited. the role of socialization and symbolic trajectories. *European Journal of Sport & Society*, *1*(1), 35-49.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., ... Edmunds, W. J. (2008, 03). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, *5*(3), e74. doi: 10.1371/ journal.pmed.0050074

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford, UK: Oxford University Pressx.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*, 026113. doi: 10.1103/PhysRevE.69.026113

Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J., & Korhonen, I. (2006, Jan). Activity classification using realistic data from wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, *10*(1), 119-128. doi: 10.1109/TITB.2005.856863

Plaza, B. (2009). Monitoring web traffic source effectiveness with google analytics an experiment with time series. *Aslib Proceedings: New Information Perspectives*, *61*(5), 474-482. doi: 10.1108/00012530910989625

Plaza, B. (2011). Google analytics for measuring website performance. *Tourism Management*, *32*(3), 477-481. doi: 10.1016/j.tourman.2010.03.015

Ridings, C., & Gefen, D. (2004). Virtual community attraction: Why people hang out online. *Journal of Computer-Mediated Communication*, *10*(1). doi: 10.1111/j.1083-6101 .2004.tb00229.x

Riemer, B. A., & Visio, M. (2003). Gender typing of sports: An investigation of metheny's classification. *Research quarterly for exercise and sport*, *74*(2), 193–204. Retrieved from `https://www.scopus.com/inward/ record.url?eid=2-s2.0-0041851011&partnerID= 40&md5=6a6307c53766f6e258223c1697c7fe85`

Russell, J. (2005). The value of dangerous sport. *Journal of the Philosophy of Sport*, *32*(1), 1-19. doi: 10.1080/00948705 .2005.9714667

Scheerder, J., Vanreusel, B., & Taks, M. (2005). Stratification patterns of active sport involvement among adults:

Social change and persistence. *International Review for the Sociology of Sport*, *40*(2), 139-162. doi: 10.1177/ 1012690205057191

Scott, J. (2013). *Social network analysis* (Third ed.). Oxford, UK: Sage Publications.

Scotton, C. (2004). Verso una classificazione definitiva delle specialit sportive. [towards an ultimate classification of sport specialities, original in italian]. *New Athletics Research in Sport Sciences*, *XXXI*(185), 26–36.

Stamm, H., & Lamprecht, M. (2005). Structural and cultural factors influencing physical activity in switzerland. *Journal of Public Health*, *13*(4), 203-211. doi: 10.1007/s10389-005 -0117-y

Starosta, W. (1987a). Alcuni problemi della tecnica sportiva [A number of sportive technique problems, original in polish]. *Scuola dello Sport*, *VI*(9), 41-44.

Starosta, W. (1987b). Alcuni problemi della tecnica sportiva [A number of sportive technique problems, original in polish]. *Scuola dello Sport*, *VI*(10), 21-27.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57*(2), pp. 307-333. Retrieved from `http://www.jstor.org/stable/ 1912557`

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge: Cambridge University Press.

Whitehead, J. (1993, March). Why children choose to do sport or stop. In M. Lee (Ed.), *Coaching children in sport-principles and practice* (p. 328). Routledge.

Wicker, P., Breuer, C., & Pawlowski, T. (2009). Promoting sport for all to age-specific target groups: The impact of sport infrastructure. *European Sport Management Quarterly*, *9*(2), 103-118.

Zar, J. H. (2010). *Biostatistical analysis* (Fifth ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Zhou, T., Ren, J., Medo, M., & Zhang, Y.-C. (2007). Bipartite network projection and personal recommendation. *Physical Review E*, *76*, 046115. doi: 10.1103/PhysRevE.76.046115