The following network of Olympic athletes

Jorge P. Rodríguez[1,2], and Lluís Arola-Fernández[2]
[1]CA UNED Illes Balears, Universidad Nacional de Educación a Distancia, Palma, Spain
[2]Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC), CSIC-UIB, Palma, Spain

Professional sport is an activity that enhances social interaction among the athletes through meetings such as training groups, large-scale competitions and sponsored events. Within these, the Olympic games represent the largest international competition, having a unique global impact. This event is assumed to promote international cooperation through the values of sport, becoming an opportunity for creating global connections between athletes.

Here we want to understand the properties of the following activity, an indicator of the information flows, between highly successful Olympic athletes both at microscopic and macroscopic scales. Thus, we built a database including the Twitter usernames of Olympic medalists in Tokyo 2020, using it for creating the follower-followee network, a directed network representing who is following whom in this social network. From our database of 1061 athletes, we found 7326 connections between 964 athletes. The most popular athletes, those with highest in-degree (*i.e.*, number of followers), were Kevin Durant (basketball, USA), Allyson Felix (athletics, USA), Teddy Riner (judo, France), Alex Morgan (football, USA) and Simone Biles (gymnastics, USA). The macro-scale network structure displayed organized patterns by sports and countries and microscopic –node– properties showed a higher activity among female athletes, with a strong sex assortativity. We quantified these properties comparing them with randomly expected values keeping fixed either the in- and out- degrees or the group sizes, revealing the underlying assortativity patterns (Fig. 1). Finally, we performed a link prediction task, quantifying the importance of follower and followee's metadata to predict the existence of a link between any two athletes.

Our network describes the following activity, which captures the interest of athletes besides real acquaintance or friendship, as well as it shows the information flows, which can be crucial for the inference of influential spreaders. We suggest that datasets built thanks to the high public exposure of the professional sport community can serve as a proxy to investigate interesting aspects of many complex socio-cultural systems at different scales.

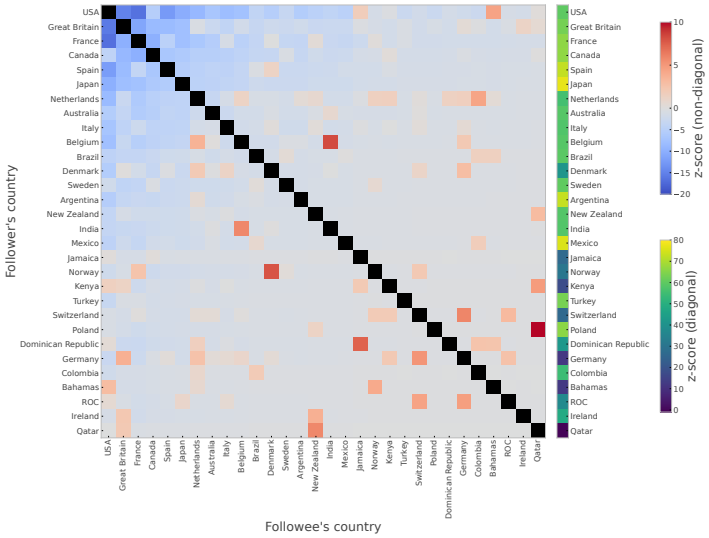

Fig. 1. Normalized following patterns among the top-30 Olympic countries with highest in-degree. Colors represent the z-score obtained comparing the observed number of following links from i to j, E^o_{ij}, with the statistics of 10^4 shufflings of the athlete's networks keeping the in- and out-degrees, average links E^s_{ij} and standard deviation σ^s_{ij}, leading to $z_{ij} = \frac{E^o_{ij} - E^s_{ij}}{\sigma^s_{ij}}$. The diagonal interactions have a different scale, so they are represented in the lateral bar.