A network approach to investigate the aggregation phenomena in sports

Luca Ferreri[1,2] Fabio Daolio[4] Marco Ivaldi[3]
Mario Giacobini[1,2] Marco Tomassini[4]

[1]Complex System Unit, Molecular Biotechnology Center
[2]Department of Animal Production, Epidemiology and Ecology, Faculty of Veterinary Medicine
[3]Motor Science Research Center, S.U.I.S.M
University of Torino, Italy
[4]Faculty of Business and Economics, Department of Information Systems
University of Lausanne, Switzerland

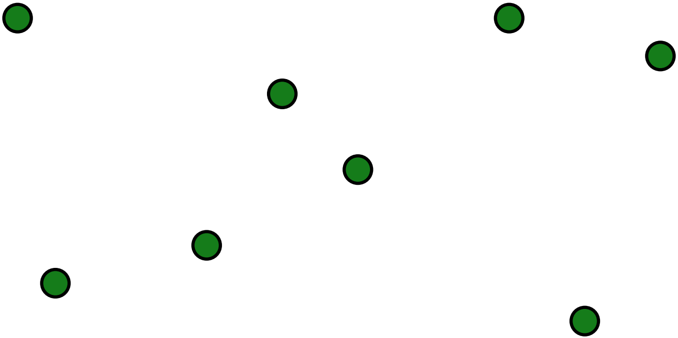
CS-SPORTS

Paris 12th August 2011

unito.it We-Sport .com We do it together Unil

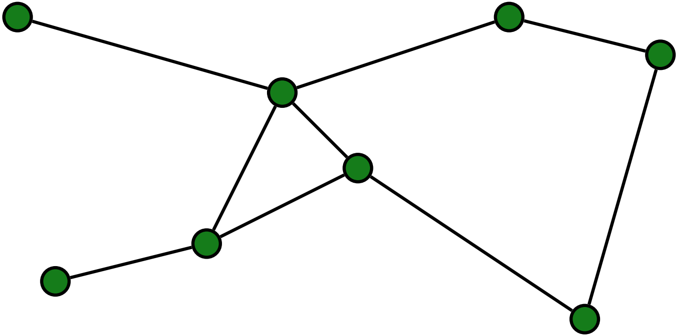
Network: definition

A network is given by a set of nodes

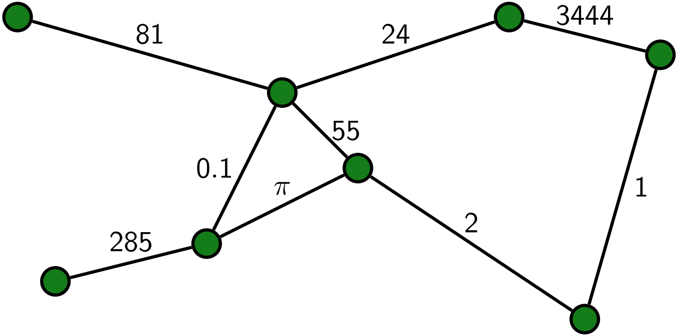


Network: definition

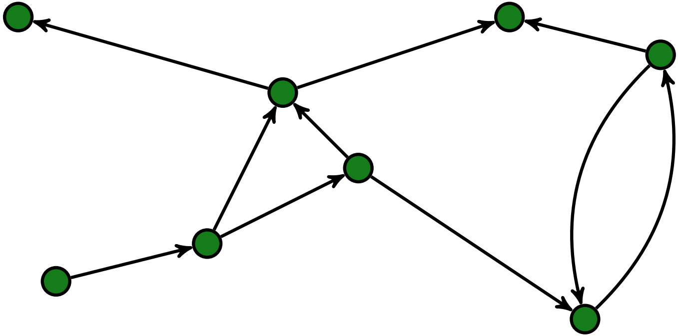
A network is given by a set of nodes and of interactions among nodes called edges



Weighted Networks



Directed Networks



Complex Networks: Nodes Centralities

- degree-centrality: the importance of a node grows proportionally with its degree

Complex Networks: Nodes Centralities

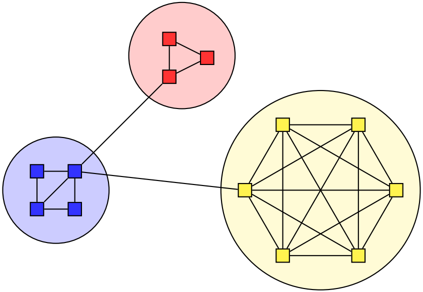
- degree-centrality: the importance of a node grows proportionally with its degree
- betweenness-centrality: the importance of a node given by the number of paths of minimum lenght that cross the node

Complex Networks: Nodes Centralities

- degree-centrality: the importance of a node grows proportionally with its degree
- betweenness-centrality: the importance of a node given by the number of paths of minimum lenght that cross the node
- eigenvector-centrality: the importance of a node is proportional to the sum of the importance of all vertices that point to it (Newman 2003):

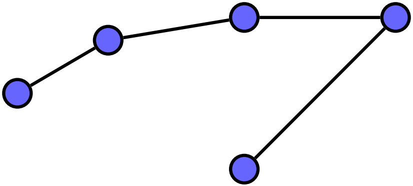
Complex Networks: Communities Detection

A community is defined as a subnet having few number of edges departing from it



Complex Networks: Distance among Nodes

Distance among nodes is defined as the minimum number of edges necessary to connect two nodes.



The shortest path in network is called the radius of the network while the longest is the diameter. In real world network it has been observed the small-world phenomena: a small diameter compared with the number of nodes.

Complex Networks: Assortativity

We try to answer the question whether nodes prefer to connect with their similar (assortative behaviour) or not (dissasortative). In particular for node similarity we intend degree similarity.

Complex Networks: Assortativity

We try to answer the question whether nodes prefer to connect with their similar (assortative behaviour) or not (dissasortative). In particular for node similarity we intend degree similarity.
Two approaches are known to detect the assortativity:

Complex Networks: Assortativity

We try to answer the question whether nodes prefer to connect with their similar (assortative behaviour) or not (dissasortative). In particular for node similarity we intend degree similarity.
Two approaches are known to detect the assortativity:

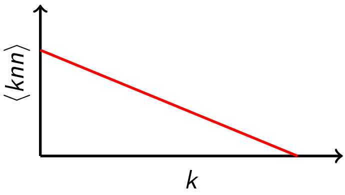
- the study of the Pearson assortative coefficient, that detect the correlation among nodes;
- the study of the average degree of the nearest neighbors

Complex Networks: Assortativity

We try to answer the question whether nodes prefer to connect with their similar (assortative behaviour) or not (dissasortative). In particular for node similarity we intend degree similarity.

Two approaches are known to detect the assortativity:

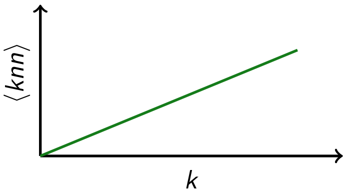
- the study of the Pearson assortative coefficient, that detect the correlation among nodes;
- the study of the average degree of the nearest neighbors



Complex Networks: Assortativity

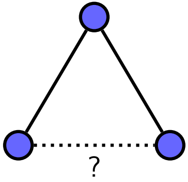
We try to answer the question whether nodes prefer to connect with their similar (assortative behaviour) or not (dissasortative). In particular for node similarity we intend degree similarity.
Two approaches are known to detect the assortativity:

- the study of the Pearson assortative coefficient, that detect the correlation among nodes;
- the study of the average degree of the nearest neighbors



Complex Networks: Clustering Coefficient

The clustering coefficient of a node is a measure of how its neighbors are connected.



Complex Networks: Heterogenity

The degree distribution of a network is the probability for a node to have a number of edges departing from it.
Complex networks could be distinct in:

Complex Networks: Heterogenity

The degree distribution of a network is the probability for a node to have a number of edges departing from it.
Complex networks could be distinct in:

- regular: having homogeneous degree distribution such as **fixed, binomial, Poisson, exponential or normal**;
- scale-free: having fat tailed degree distribution well described by **power law** distribution (at least asymptotically)

Complex Networks: Heterogenity

The degree distribution of a network is the probability for a node to have a number of edges departing from it.
Complex networks could be distinct in:

- regular: having homogeneous degree distribution such as **fixed, binomial, Poisson, exponential or normal**;
- scale-free: having fat tailed degree distribution well described by **power law** distribution (at least asymptotically);

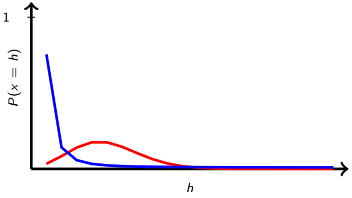
a measure of the heterogeneity is given by:

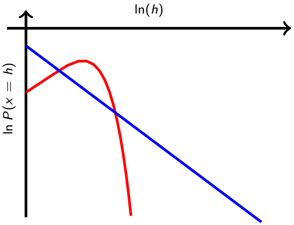
$$\frac{\langle k \rangle}{\langle k^2 \rangle}$$

Complex Networks: Heterogenity

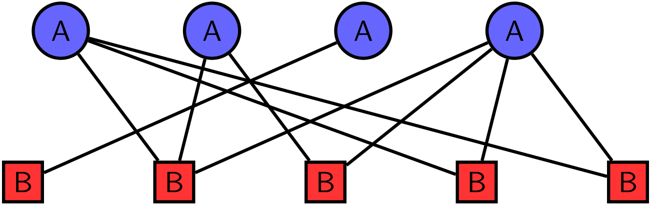
The degree distribution of a network is the probability for a node to have a number of edges departing from it.
Complex networks could be distinct in:

- regular: having homogeneous degree distribution such as **fixed, binomial, Poisson, exponential or normal**;
- scale-free: having fat tailed degree distribution well described by **power law** distribution (at least asymptotically);





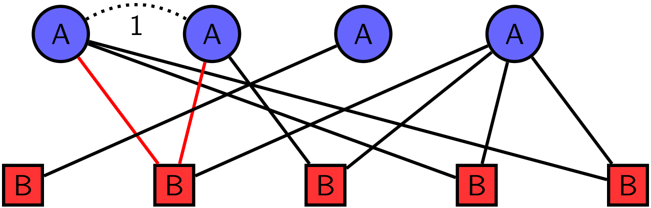
Bipartite Networks



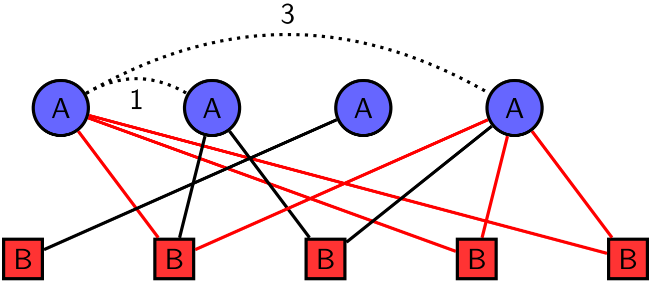
Many example:

- co-authorship network;
- diseasome;
- heterosexual contact network;
- vector-borne disease network;

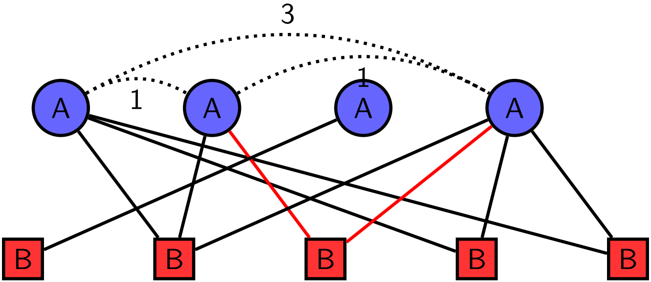
Bipartite Networks



Bipartite Networks

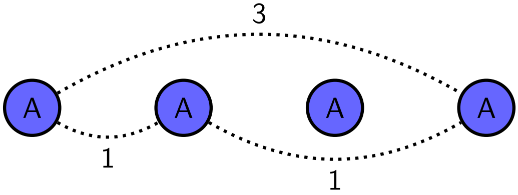


Bipartite Networks

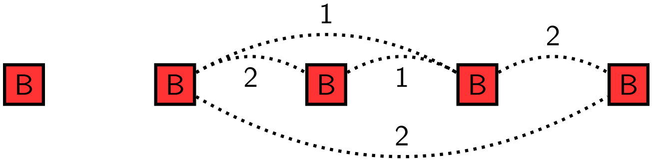


Bipartite Networks

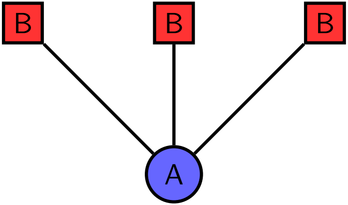
The A-projection



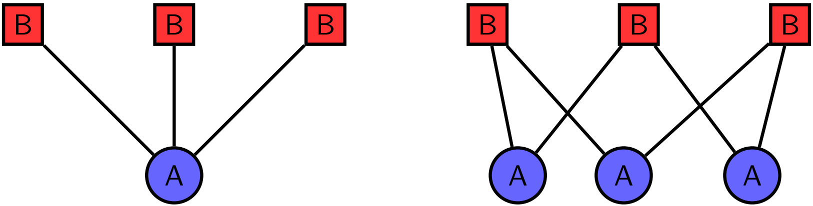
the B-projection



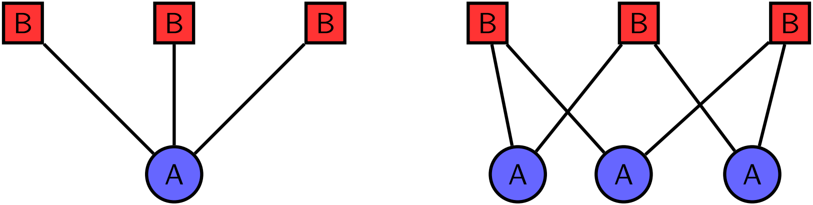
Bipartite Projection is less informative



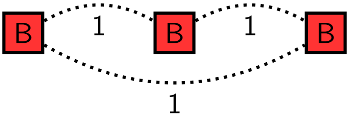
Bipartite Projection is less informative



Bipartite Projection is less informative



are both projected in



We-Sport: a sparse network

We consider a snapshot of the entire network:

- 1680 athletes

We-Sport: a sparse network

We consider a snapshot of the entire network:

- 1680 athletes
- 240 sports

We-Sport: a sparse network

We consider a snapshot of the entire network:

- 1680 athletes
- 240 sports
- 6107 interactions

We-Sport: a sparse network

We consider a snapshot of the entire network:

- 1680 athletes
- 240 sports
- 6107 interactions

we define the density for bipartite network as:

$$\delta = \frac{\text{edges}}{\text{sports} \cdot \text{athletes}} \simeq 0.014$$

We-Sport: a sparse network

We consider a snapshot of the entire network:

- 1680 athletes
- 240 sports
- 6107 interactions

we define the density for bipartite network as:

$$\delta = \frac{\text{edges}}{\text{sports} \cdot \text{athletes}} \simeq 0.014$$

We observe only two large connected components:

- the first have 1679 athletes and 239 sports

We-Sport: a sparse network

We consider a snapshot of the entire network:

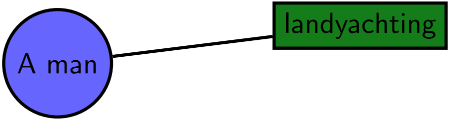
- 1680 athletes
- 240 sports
- 6107 interactions

we define the density for bipartite network as:

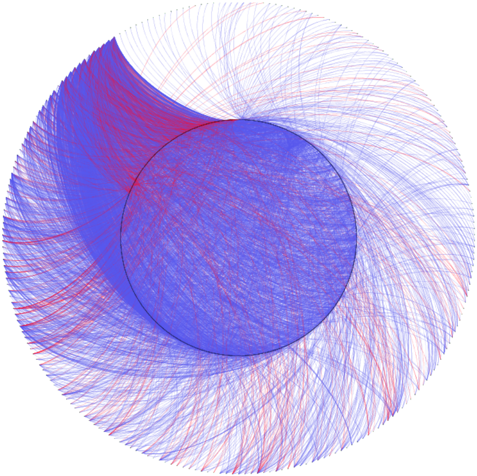
$$\delta = \frac{\text{edges}}{\text{sports} \cdot \text{athletes}} \simeq 0.014$$

We observe only two large connected components:

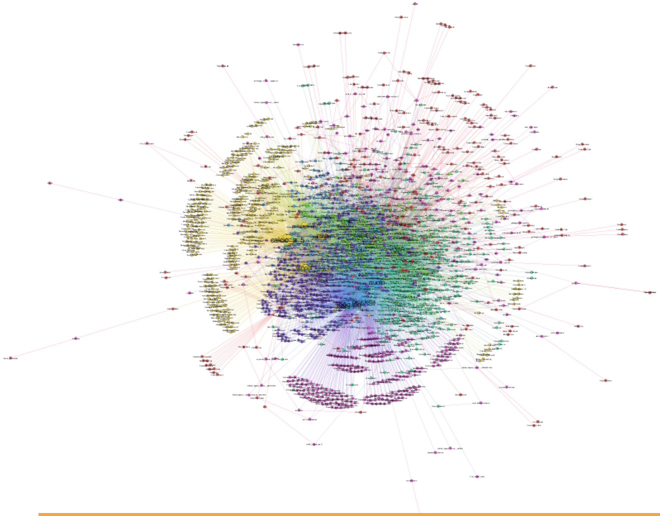
- the first have 1679 athletes and 239 sports
- the second



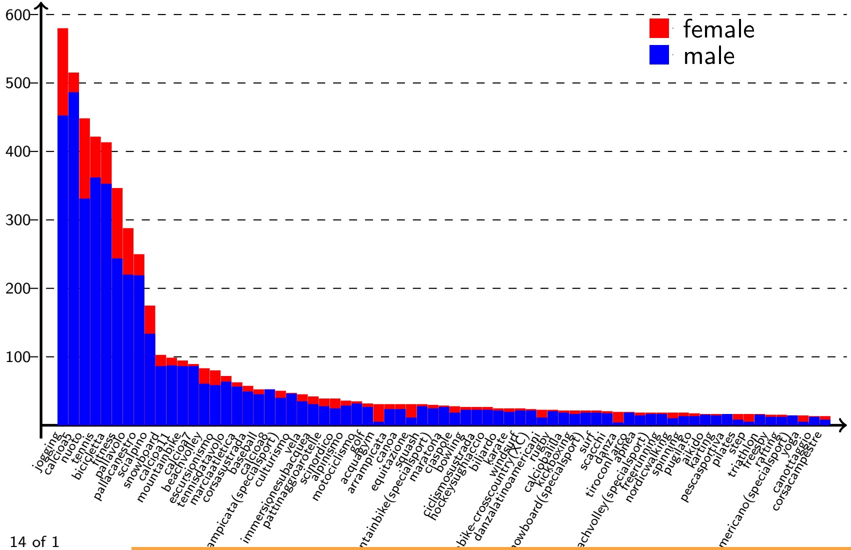
Graphical representation of bipartite We-Sport network



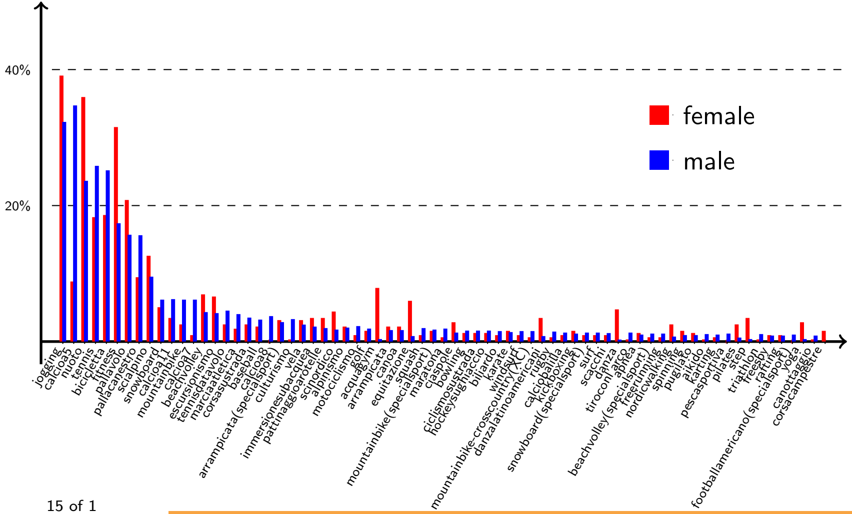
Graphical representation of bipartite We-Sport network



The 70 most played Sports



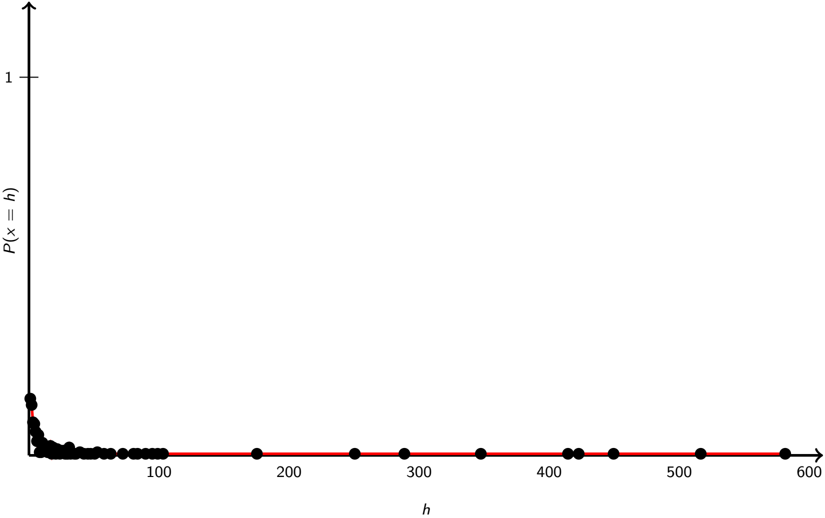
The 70 most played Sports: gender frequencies



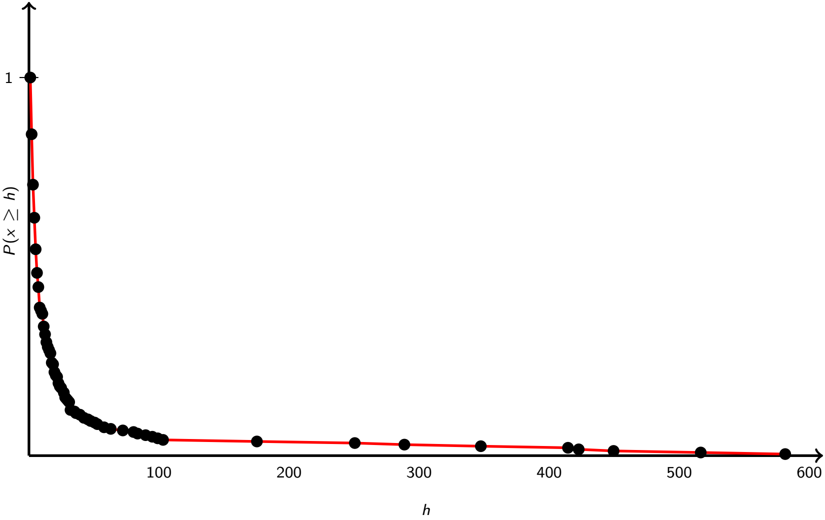
A complex network

partition	mode	median	$\langle k \rangle$	$\langle k^2 \rangle$	$\frac{\langle k \rangle}{\langle k^2 \rangle}$
sport	1	5	25.54	$6.183 \cdot 10^3$	0.0041
athletes	1	3	3.63	23.78	0.1530

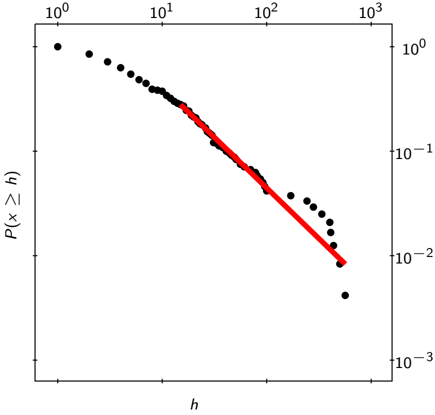
Degree distribution: sport nodes



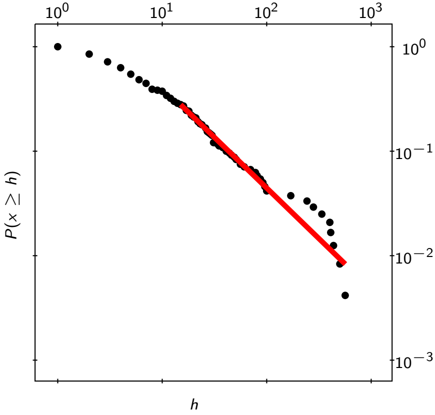
Degree distribution: sport nodes



Degree distribution: sport nodes logarithmic scale

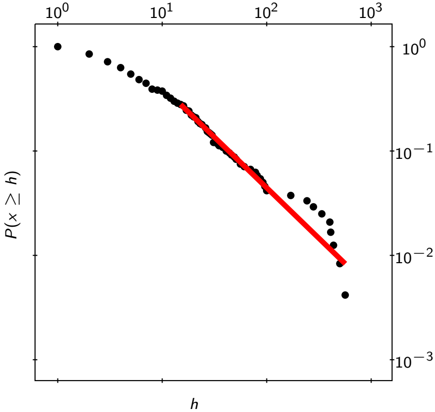


Degree distribution: sport nodes logarithmic scale



$\alpha = 1.96$
$x_{\text{min}} = 15$
p-value$=$ 0.5670

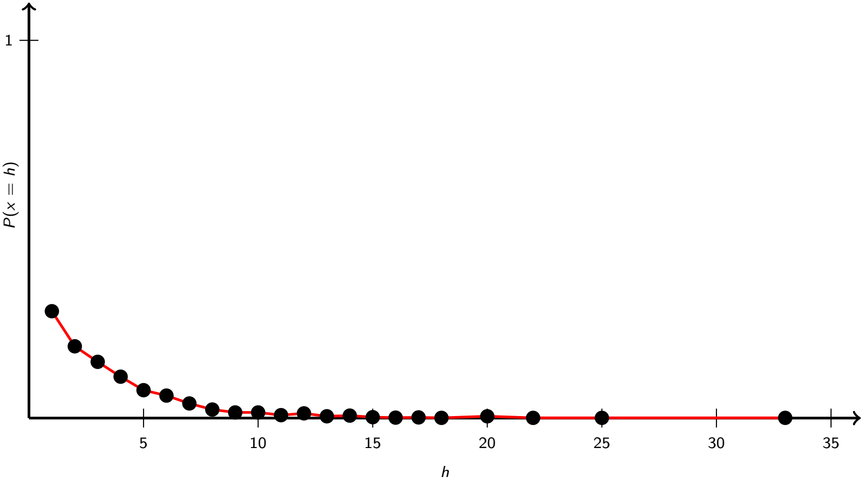
Degree distribution: sport nodes logarithmic scale



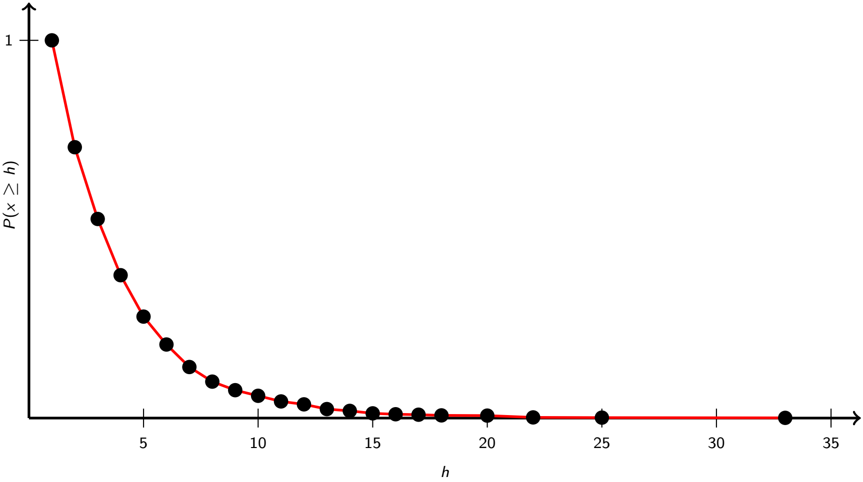
$\alpha = 1.96$
$x_{\text{min}} = 15$
p-value$= 0.5670$

	bin.neg.		Poisson		exponential		Weibull		log-normal		Yule		power law + cutoff	
	LR	p	LR	p	LR	p	LR	p	LR	p	LR	p	LR	p
sport - nodes	4.14	**0.00**	4.09	**0.00**	4.31	**0.00**	0.13	0.89	−0.35	0.72	−0.004	0.94	−0.53	0.30

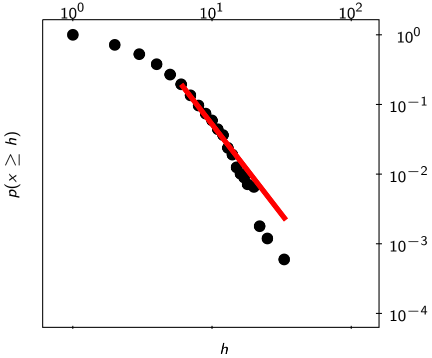
Degree distribution: athletes nodes



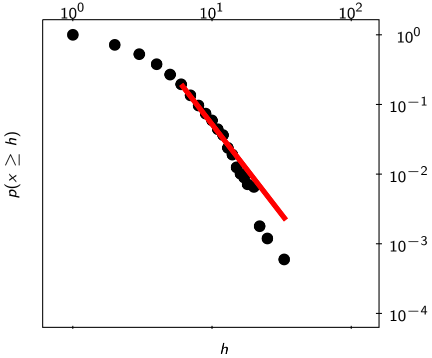
Degree distribution: athletes nodes



Degree distribution: athletes nodes logarithmic scale

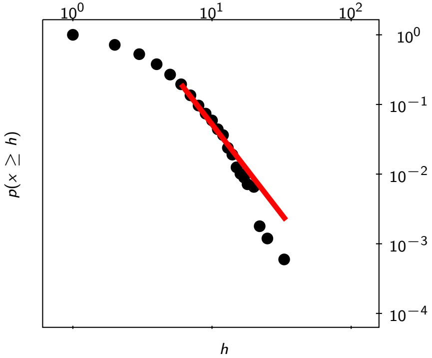


Degree distribution: athletes nodes logarithmic scale



$\alpha = 3.49$
$x_{\text{min}} = 6$
p-value$= 0.0730$

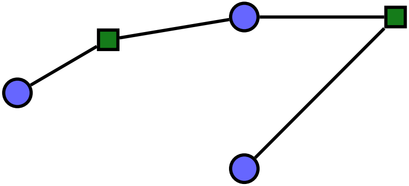
Degree distribution: athletes nodes logarithmic scale



$\alpha = 3.49$
$x_{\text{min}} = 6$
p-value$= 0.0730$

	bin.neg.		Poisson		exponential		Weibull		log-normal		Yule		power law + cutoff	
	LR	p	LR	p	LR	p	LR	p	LR	p	LR	p	LR	p
athletes - nodes	3.73	**0.00**	3.09	**0.00**	−0.83	0.40	−2.16	**0.03**	−2.12	**0.03**	−3.94	**0.00**	−6.62	**0.00**

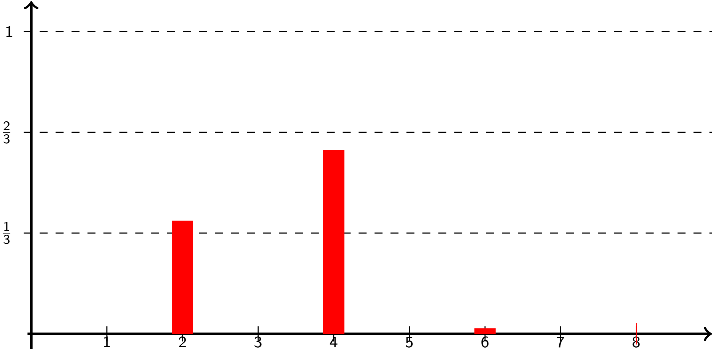
Nodes distance



The maximum distance between every pairs of nodes in a graph is defined as the diameter of the graph. We observe a diameter of 8 but on average the shortest path between nodes is 3.33.
We are in presence of so called small-world phenomena.

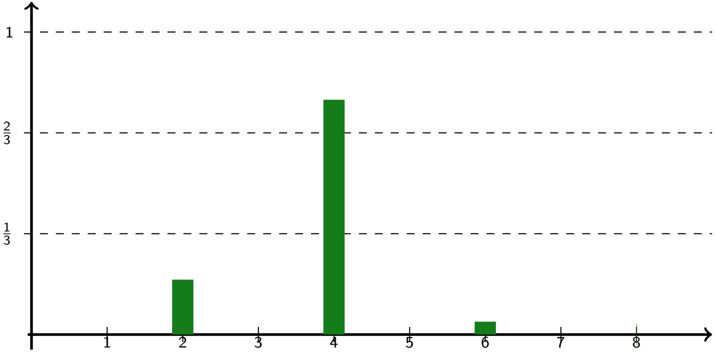
The distance distribution

the athletes-athletes distance



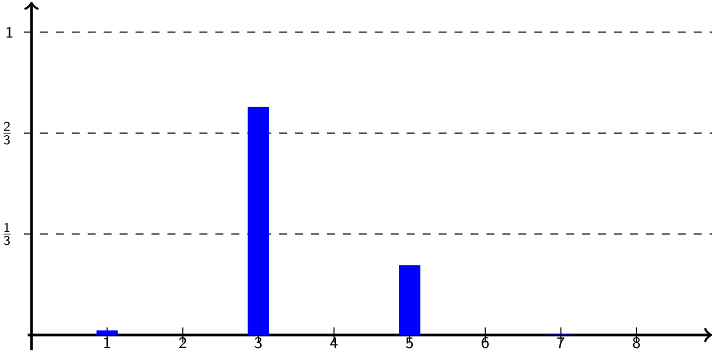
The distance distribution

the sport-sport distance



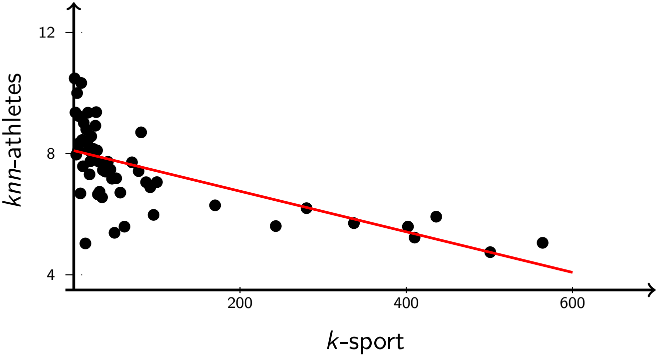
The distance distribution

the sport-athletes distance



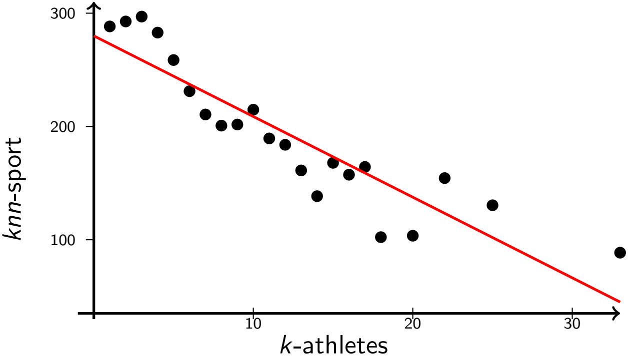
Assortativity in bipartite

We-sport network shows a disassortative behaviour: the Pearson coefficient is -0.2425.
Moreover if we calculate the nearest neighbor degree:

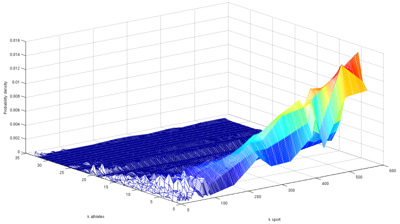


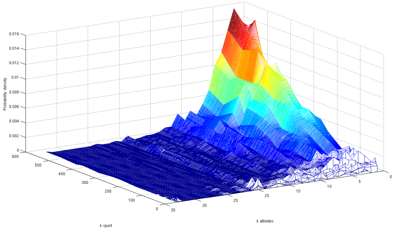
Assortativity in bipartite

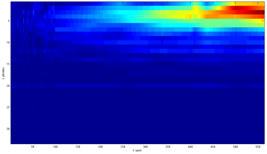
We-sport network shows a disassortative behaviour: the Pearson coefficient is -0.2425.
Moreover if we calculate the nearest neighbor degree:



The joint probability







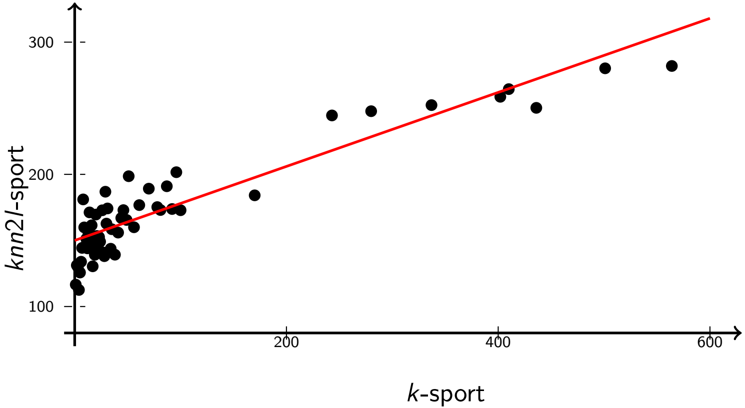
2-length assortativity in bipartite

We want to try to answer the question:

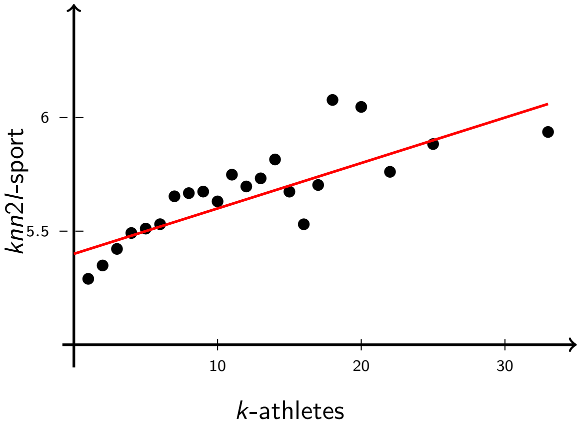
do people choose sports that connect them with similar people or not?

Therefore we analyze the 2-length assortativity: we observe a weak assortative behavior for athletes-nodes (0.0326) and stronger for sport-nodes (0.2620)

2-length assortativity in bipartite



2-length assortativity in bipartite



The Clustering Coefficient for Biparite Networks

Again in order to understand the aggregation behavior of athletes we try to understand if people prefer to connect with other sharing the same sport's preference. Hence we define a similarity matrix cc which counts for each couple of athletes the number of sports they share:

$$|N(v) \cap N(u)|$$

then we can normalize that matrix. Le Blond et al., Latapy et al., and Borgatti suggest the three following denominator:

- $\min(|N(v)|, |N(u)|)$ for $cc_{\underline{\bullet}}(u, v)$

The Clustering Coefficient for Biparite Networks

Again in order to understand the aggregation behavior of athletes we try to understand if people prefer to connect with other sharing the same sport's preference. Hence we define a similarity matrix cc which counts for each couple of athletes the number of sports they share:

$$|N(v) \cap N(u)|$$

then we can normalize that matrix. Le Blond et al., Latapy et al., and Borgatti suggest the three following denominator:

- $\min(|N(v)|, |N(u)|)$ for $cc_{\underline{\bullet}}(u, v)$
- $\max(|N(v)|, |N(u)|)$ for $cc_{\overline{\bullet}}(u, v)$

The Clustering Coefficient for Biparite Networks

Again in order to understand the aggregation behavior of athletes we try to understand if people prefer to connect with other sharing the same sport's preference. Hence we define a similarity matrix cc which counts for each couple of athletes the number of sports they share:

$$|N(v) \cap N(u)|$$

then we can normalize that matrix. Le Blond et al., Latapy et al., and Borgatti suggest the three following denominator:

- $\min(|N(v)|, |N(u)|)$ for $cc_{\underline{\bullet}}(u, v)$
- $\max(|N(v)|, |N(u)|)$ for $cc_{\overline{\bullet}}(u, v)$
- $|N(v) \cap N(u)|$ for $cc_{\bullet}(u, v)$

The clustering coefficient II

From the similarity matrix we can calculate the clustering coefficient of each nodes.

$$cc(v) = \frac{\sum_{u \in N(N(v))} cc(v, u)}{|N(N(v))|}$$

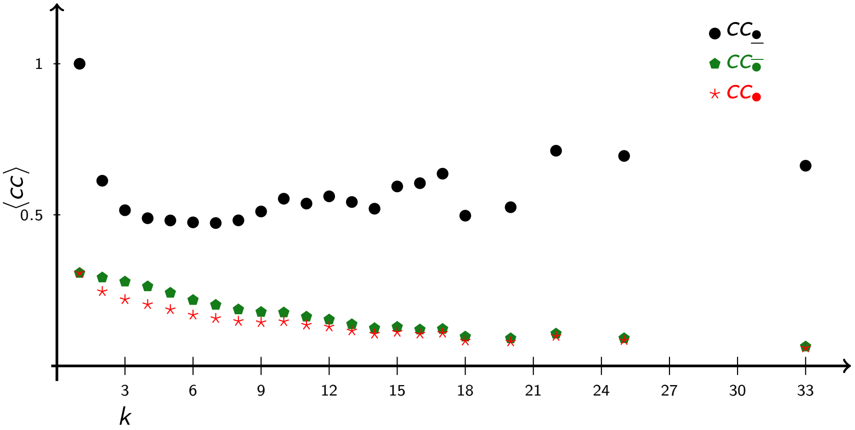
and from that the clustering coefficient of A-partition:

$$cc = \frac{1}{|A|} \sum_{v \in A} cc(v)$$

graph	$cc_{\underline{\bullet}}$	$cc_{\overline{\bullet}}$	$cc_{\bullet}$
athletes	0.6628	0.2672	0.2315
sport	0.4126	0.0615	0.0536

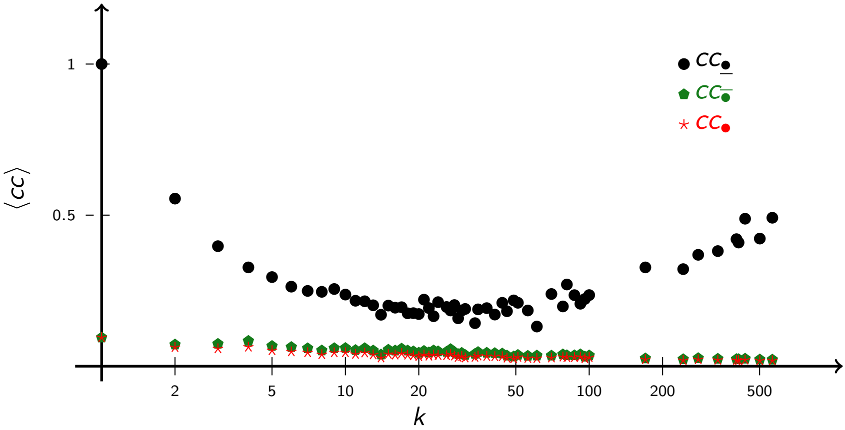
The clustering coefficient II

The athletes case

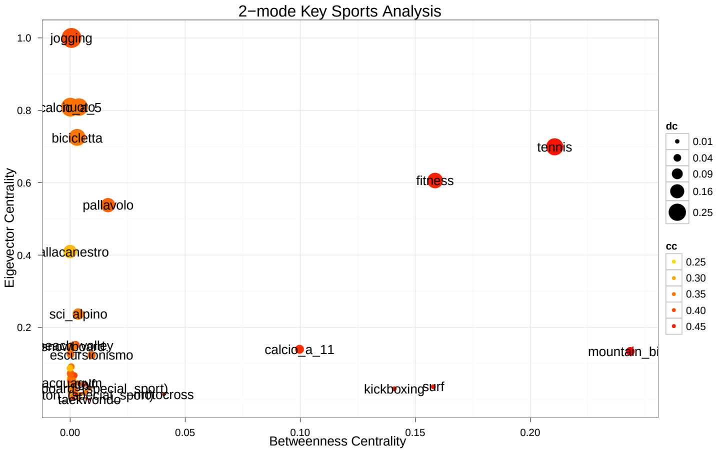


The clustering coefficient II

The sport case



The centrality



An application: which is the best sport to meet girls?

Contacts

for further informations:

www.we-sport.com

or contact us:

luca.ferreri@unito.it
fabio.daolio@unil.ch