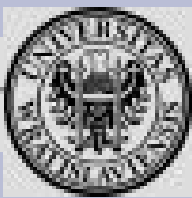




• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

**Rzeszów, 09.05.2009**

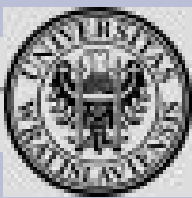


# Anomalous interactions in network of Polish Football League

*Andrzej Jarynowski  
student of 2 year Msc program: computer  
physics at University of Wrocław*



• KUSI •



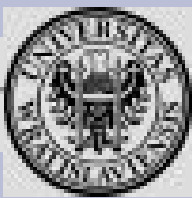
The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Program of talk

- Inspiration and history of the problem
- Hypothesis on corruption
- Introduction to seeking for anomalies
- Random matrix theory
- Nonlinear measures of dependence



✦ **KUSI** ✦



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

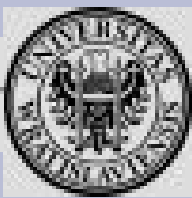
# Inspiration

## Poland-Germany





KUSI



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

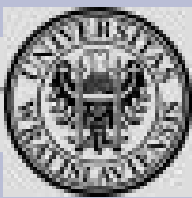
# Inspiration

- Not enough knowledge of Polish football league
- Polish reality:





+ KUSI +



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Inspiration

In recent years prosecution in Poland has been investigating several clubs, referees and players because of corruption procedures. The problem is very big. Some people said that almost everyone had been involved in corruption.

Example:

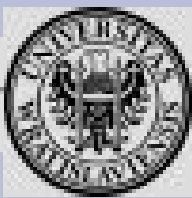
Bet companys don't allow to bet only one match in polish league-You have to combine it with at least 2 other matches.

Why?

Because some people could know who win this „fixed” match and earn money on it.



• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

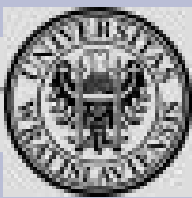
# Inspiration

We study the statistical properties of results in Polish League, looking for evidence of non-sport activity.

We treat league as a complex system and we use tools from statistical physics to research some of its properties.



• KUSI •



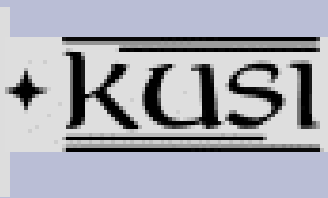
The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# History

*Prof. Bolesław Kopociński*  
(University of Wrocław)

*COMPONENTS OF THE GAME RESULT  
IN A FOOTBALL LEAGUE, 2001*

*UNFINISHED SEASON OF FOOTBALL,  
2001*



# History

Stochastic components of Kopociński model:

- the home-field advantage(3)
- the difference of the strengths of the teams(1)(2)
- a random factor(4) (5)

$$X_{ij} = \Pi^{(1)}(a_1(r_{ij})) + \Pi^{(3)}(b(m_i)) + \Pi^{(4)}(c_1)$$

$$Y_{ij} = \Pi^{(2)}(a_2(r_{ji})) + \Pi^{(5)}(c_2)$$

$a_1, a_2, b$  – are known functions

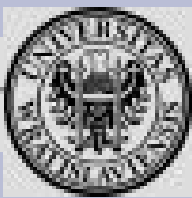
$c_1, c_2$  – parameters for league

$r, m$  – parameters for teams





• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# History

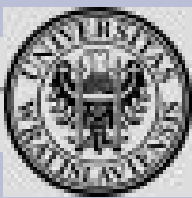
Kopociński, B. Unfinished league season of football (2001)

<b>G</b>	<b>Teams</b>	<b>Games</b>	<b>Points</b>	<b>Goals</b>
1	Ruch Chorzów	14	18	50-23
2	Wisła Kraków	12	16	31-20
3	Pogoń Lwów	13	16	27-22
4	AKS Chorzów	12	15	30-14
5	Warta Poznań	12	15	34-20
6	Cracovia	13	14	23-32
7	Polonia Warszawa	12	12	28-28
8	Garbarnia Kraków (b)	13	10	17-32
9	Warszawianka Warszawa	11	5	16-29
10	Union-Touring Łódź (b)	12	3	15-51

**Polish  
League  
End of  
August  
1939**



• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# History

Kopociński, B. Unfinished league season of football (2001)

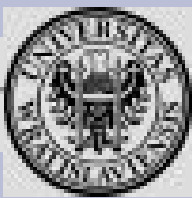
<b>G</b>	<b>Teams</b>	<b>Games</b>	<b>Points</b>	<b>Goals</b>
1	Ruch Chorzów	14	18	50-23
2	Wisła Kraków	12	16	31-20
3	Pogoń Lwów	13	16	27-22
4	AKS Chorzów	12	15	30-14
5	Warta Poznań	12	15	34-20
6	Cracovia	13	14	23-32
7	Polonia Warszawa	12	12	28-28
8	Garbarnia Kraków (b)	13	10	17-32
9	Warszawianka Warszawa	11	5	16-29
10	Union-Touring Łódź (b)	12	3	15-51

**Polish  
League  
End of  
August  
1939**

**Unfinished league season due to  
Second World War**



★ KUSI ★



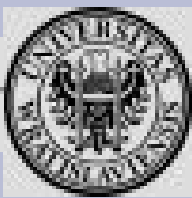
The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Hypothesis of corruption





• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

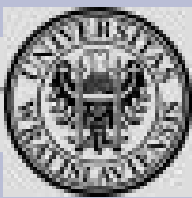
# Hypothesis of corruption

If it is true, that many of matches were „fixed”, so polish league's statistics should be different than other „sport” leagues like Bundesliga

There should be also anomalies compared to stochastic patern



\* KUSI \*



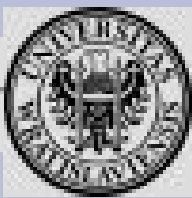
The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Hypothesis of corruption

- I consider 3 seasons of Ekstraklasa -16 teams since 2005/2006 and First League -16 teams (chosen seasons especially 2003/2004)
- I choose measure of match result:
  - 1) points (3-win, 1-draw, 0-loss)
  - 2) difference of goals in match
  - 3) "logic" (1, 0 -1)
- Focus on teams, which had fines, because of selling or buying matches, e.g. Arka Gdynia



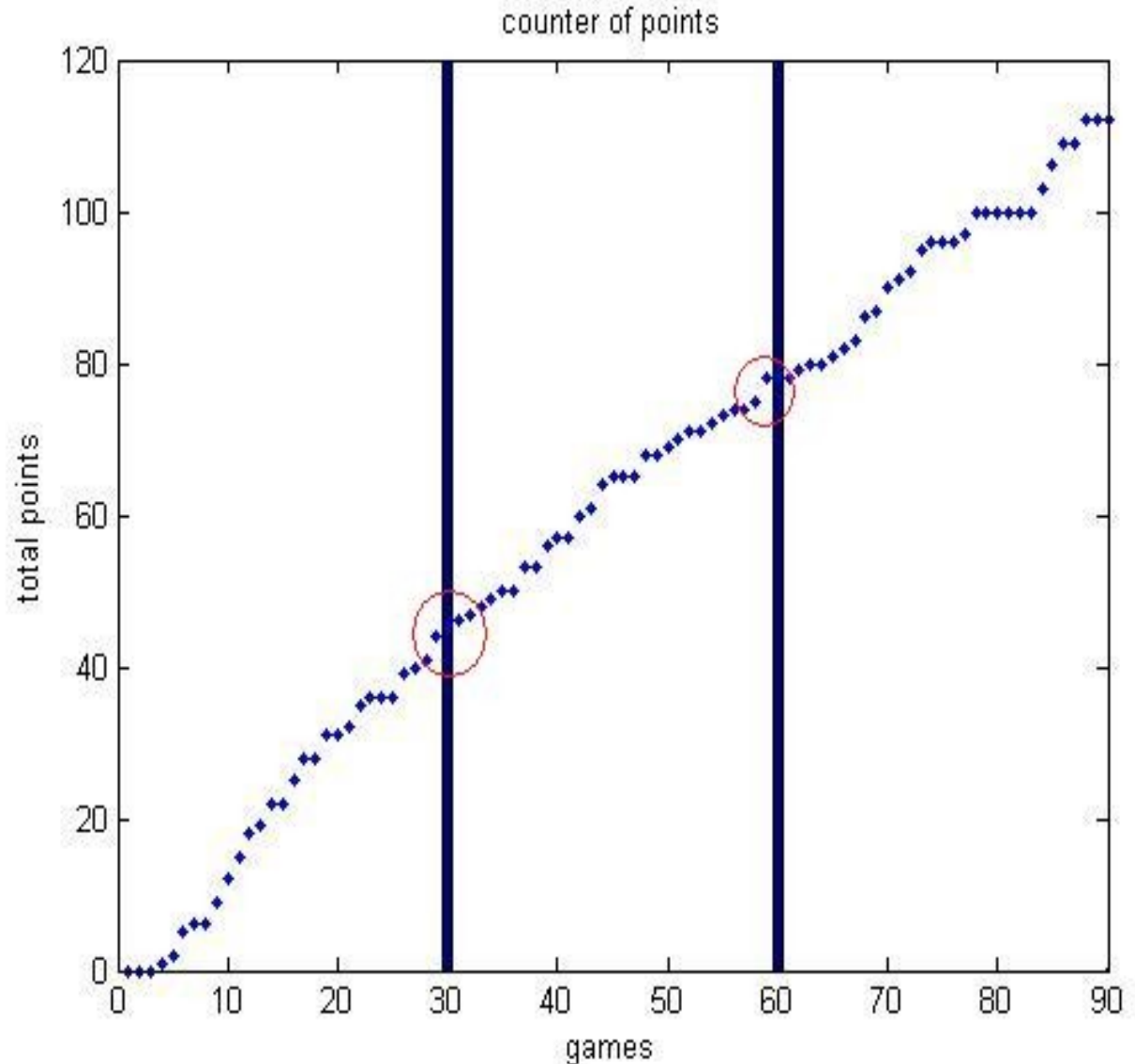
KUSI



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

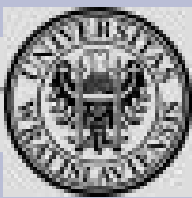
# Hypothesis on corruption

Let's consider cumulative number of points after some games.





KUŚI

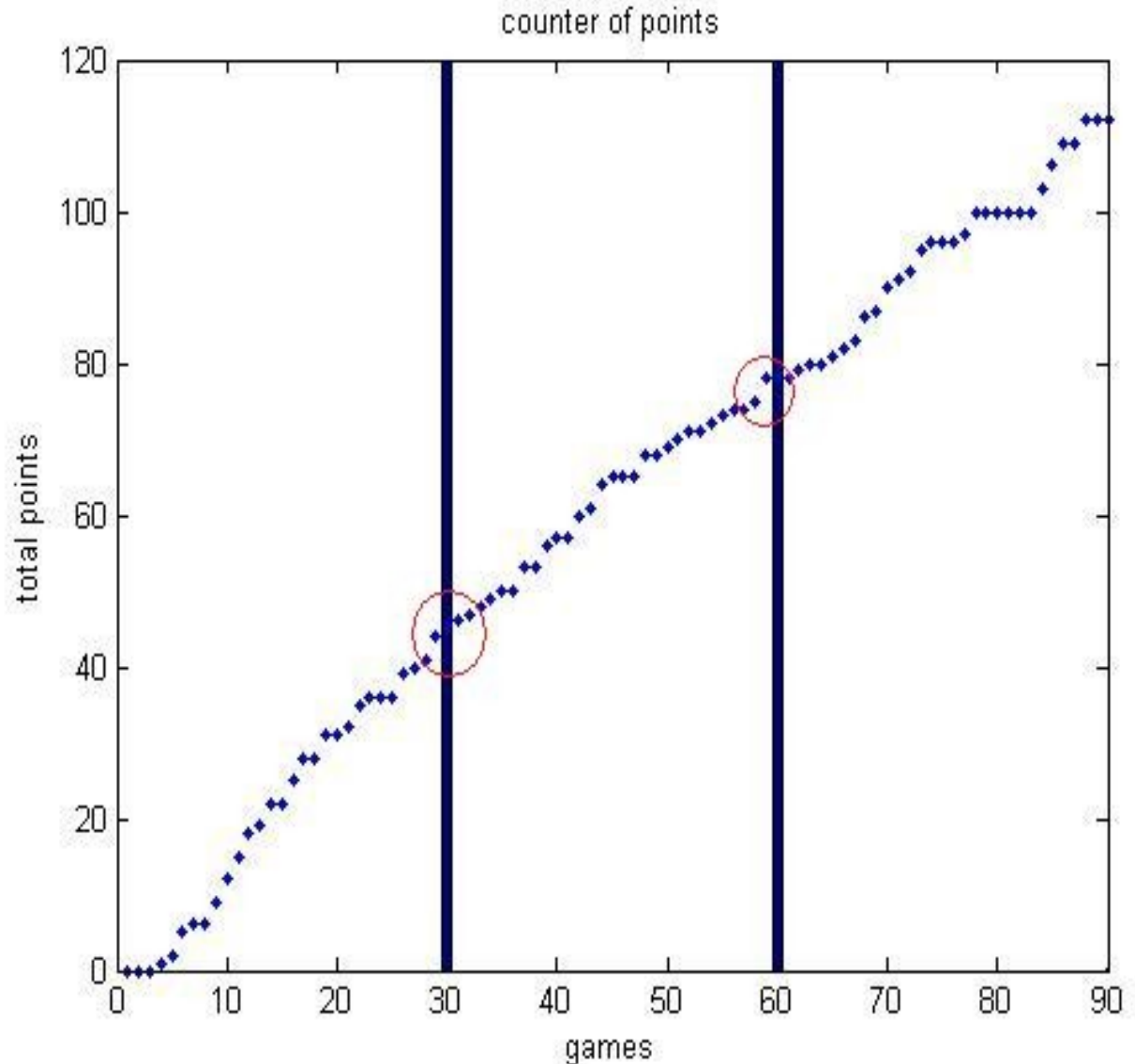


The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Hypothesis on corruption

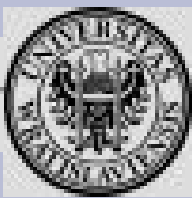
Let's consider cumulative number of points after some games.

Graph for Arka Gdynia





+ KUSI +



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Introduction to anomalies

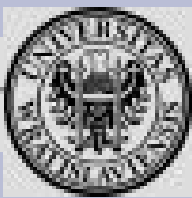
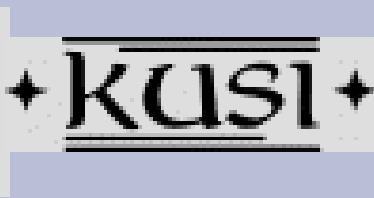
## First problems

- Problems with getting data from database in time sequence
- Short length of time series – 90
- Seasonality of data
- Changing teams (degradation and promotion)

## First solves

- 12 teams
- 90 plays – 3 seasons





# Introduction to anomalies

- Looking for linear correlation between teams
- Structure of crossing correlation coefficients for Arka

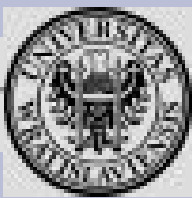
-0.0155	0.0355	0.0879	-0.0450
0.0318	-0.0096	-0.0258	0.1216
-0.1973	-0.1495	1.0000	0.0959

	zagłębie lubin	odra włodzisław	widzew	Legia	lech poznań	
korona kielce	gornik zabrze	Gks Belchatow	Dyskobolia	Cracovia	Arka Gdynia	Wisla Krakow





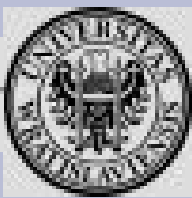
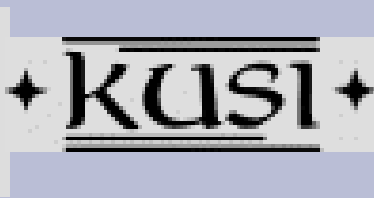
• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

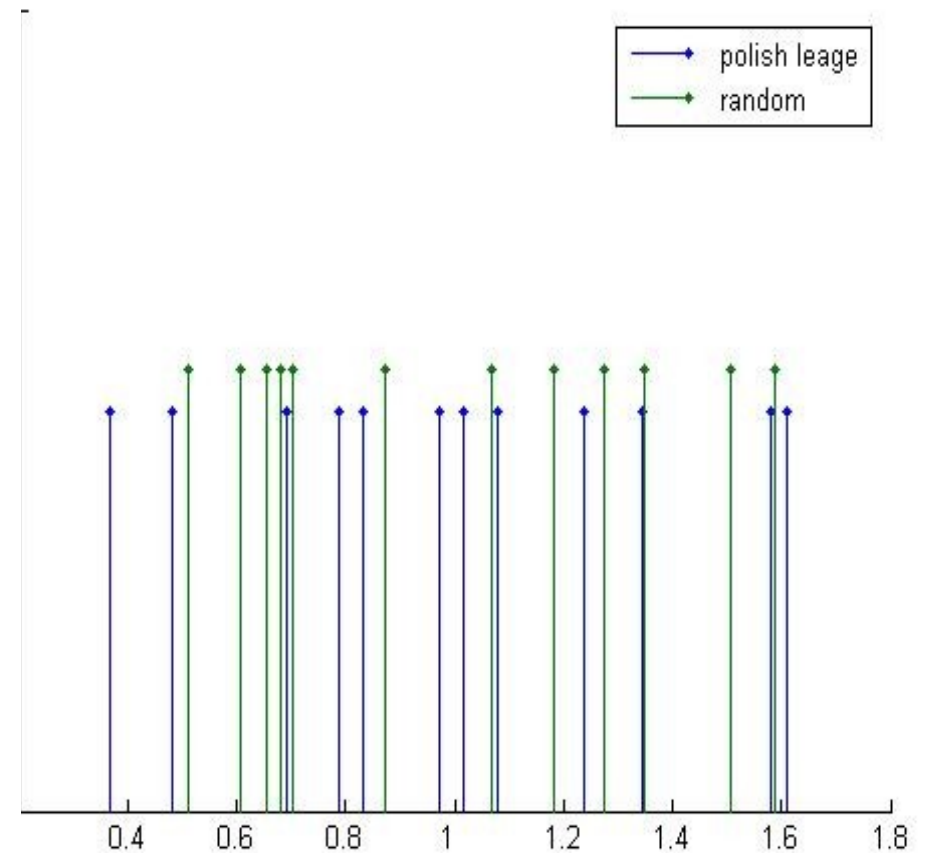
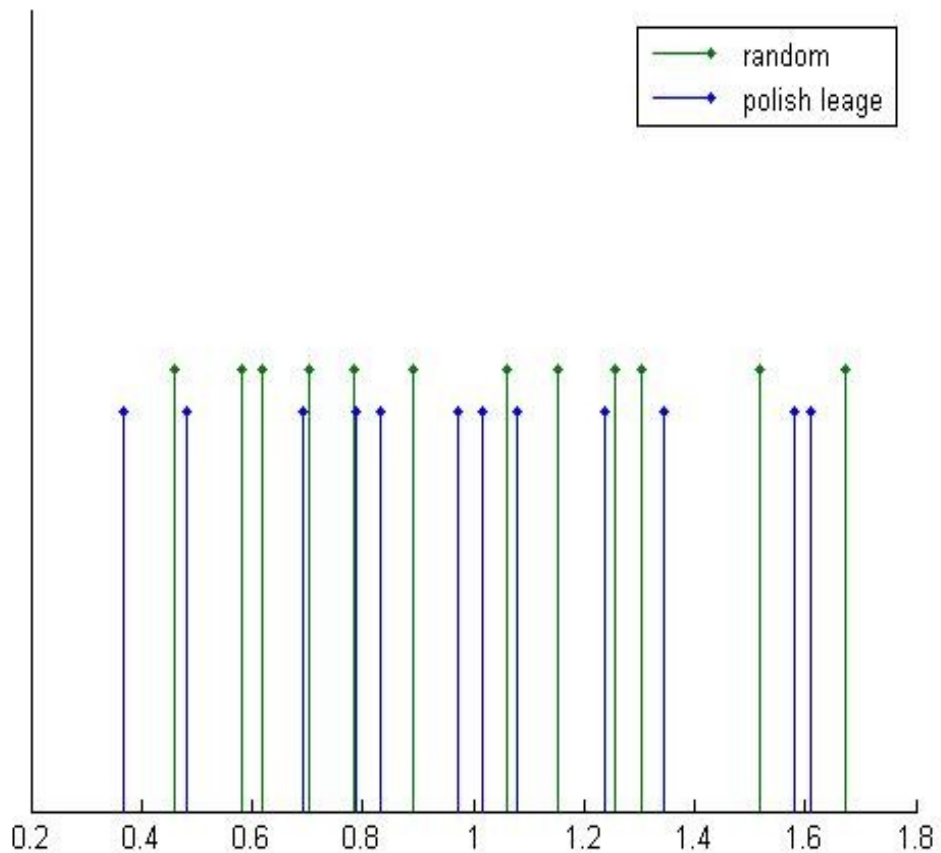
# Random Matrix Theory

- The idea is to compare spectrum of matrix of correlation to random matrix with similar statistics (in our case matrix with football data  $12 \times 90$  produce matrix of correlation  $12 \times 12$ )
- In theory we should take matrix made by gaussian noise, but empirical distribution of results is different



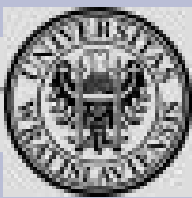
# Random Matrix Theory

Random-Unitary distribution  $(-4, 4)$  continuous-left or discrete-right  
Polish league-measure 2) difference of goals





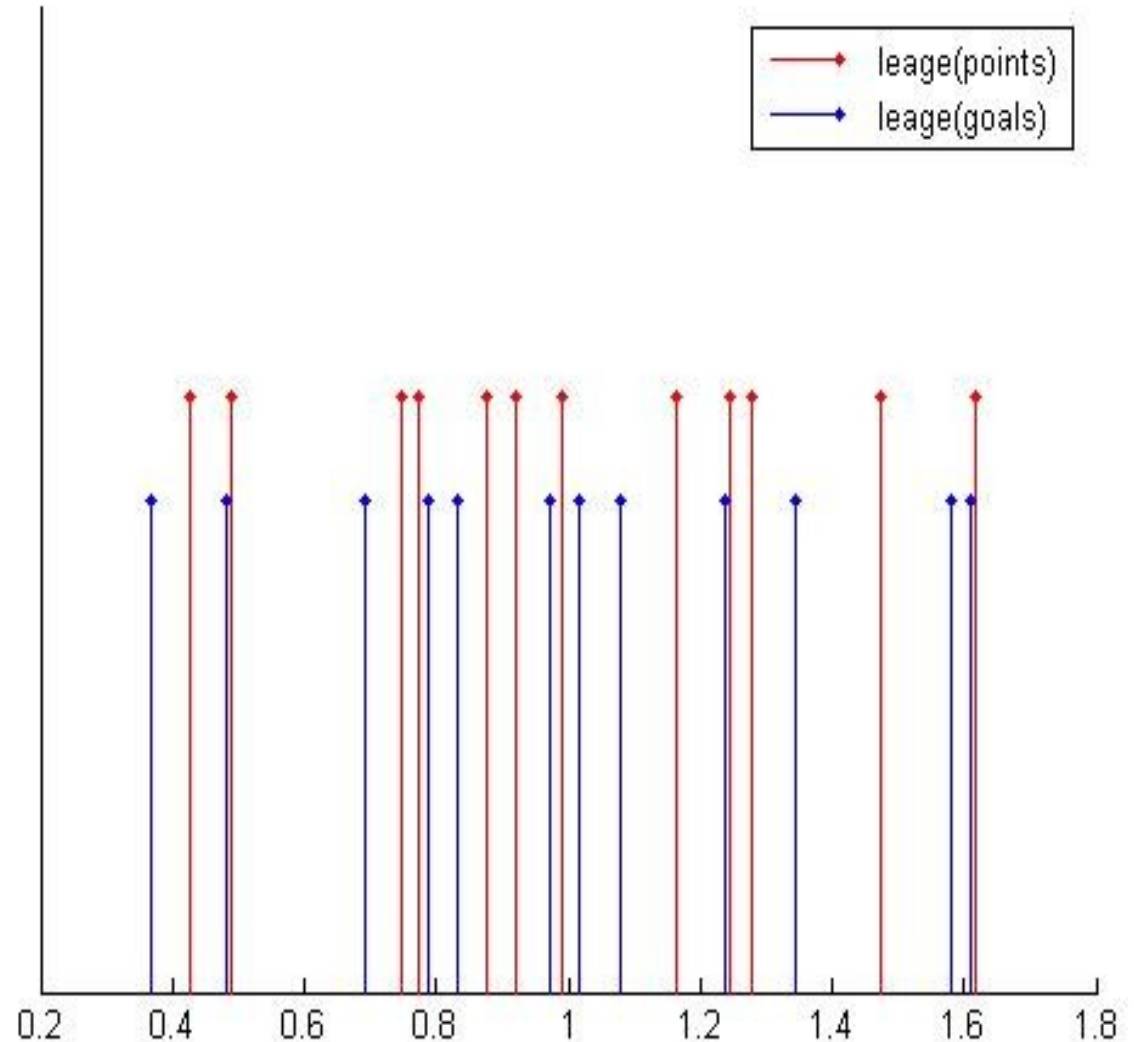
KUSI



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

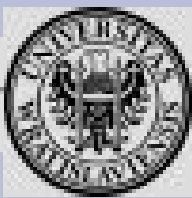
# Random Matrix Theory

league(points)-  
measure 1)  
points  
league(goals)-  
measure 2)  
difference of  
goals





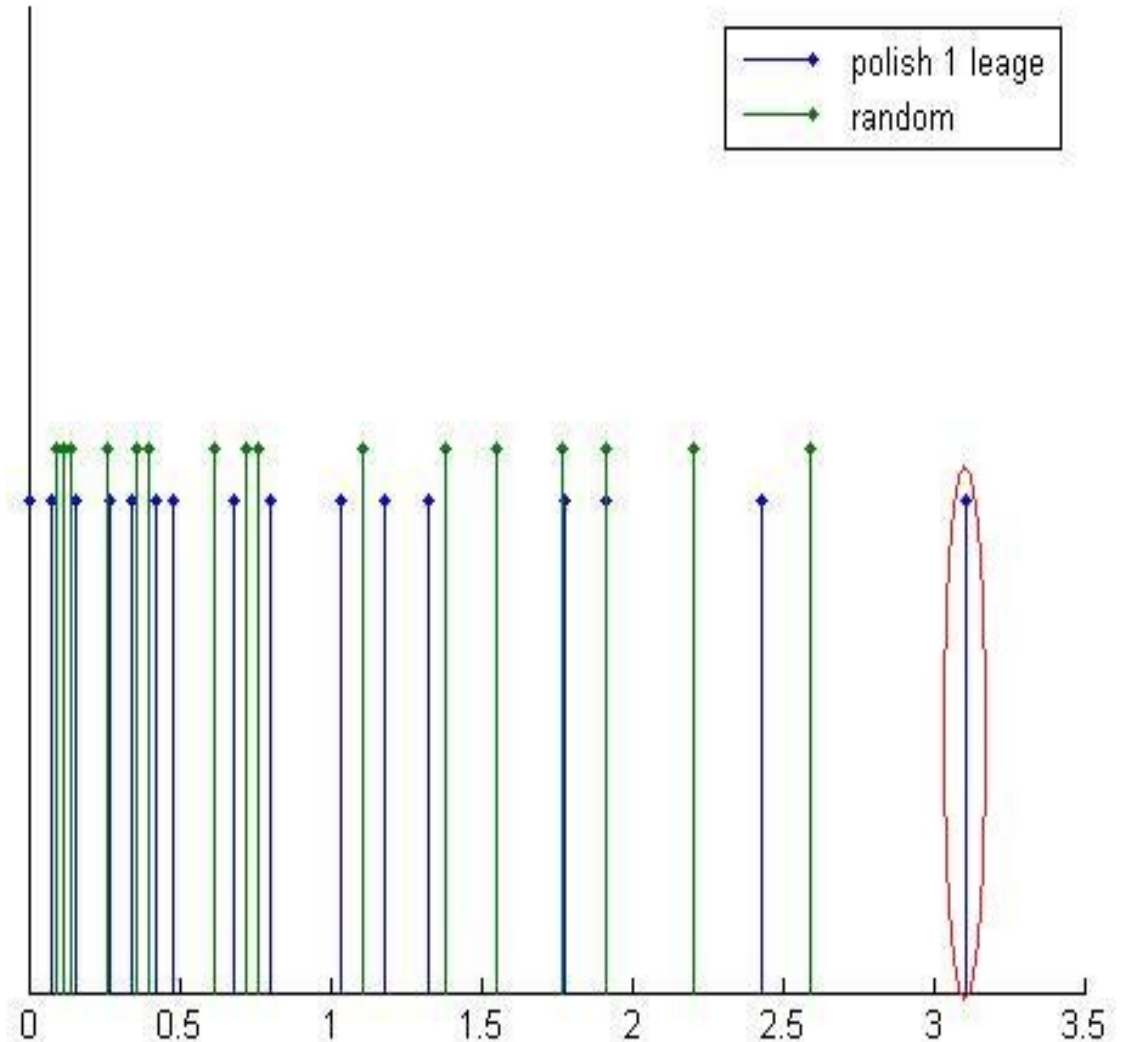
KUSI



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

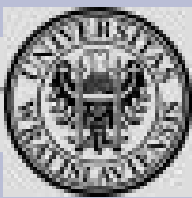
# Random Matrix Theory

Spectral analysis of  
lower polish division:  
First league  
(16teams), one  
season 2003/2004





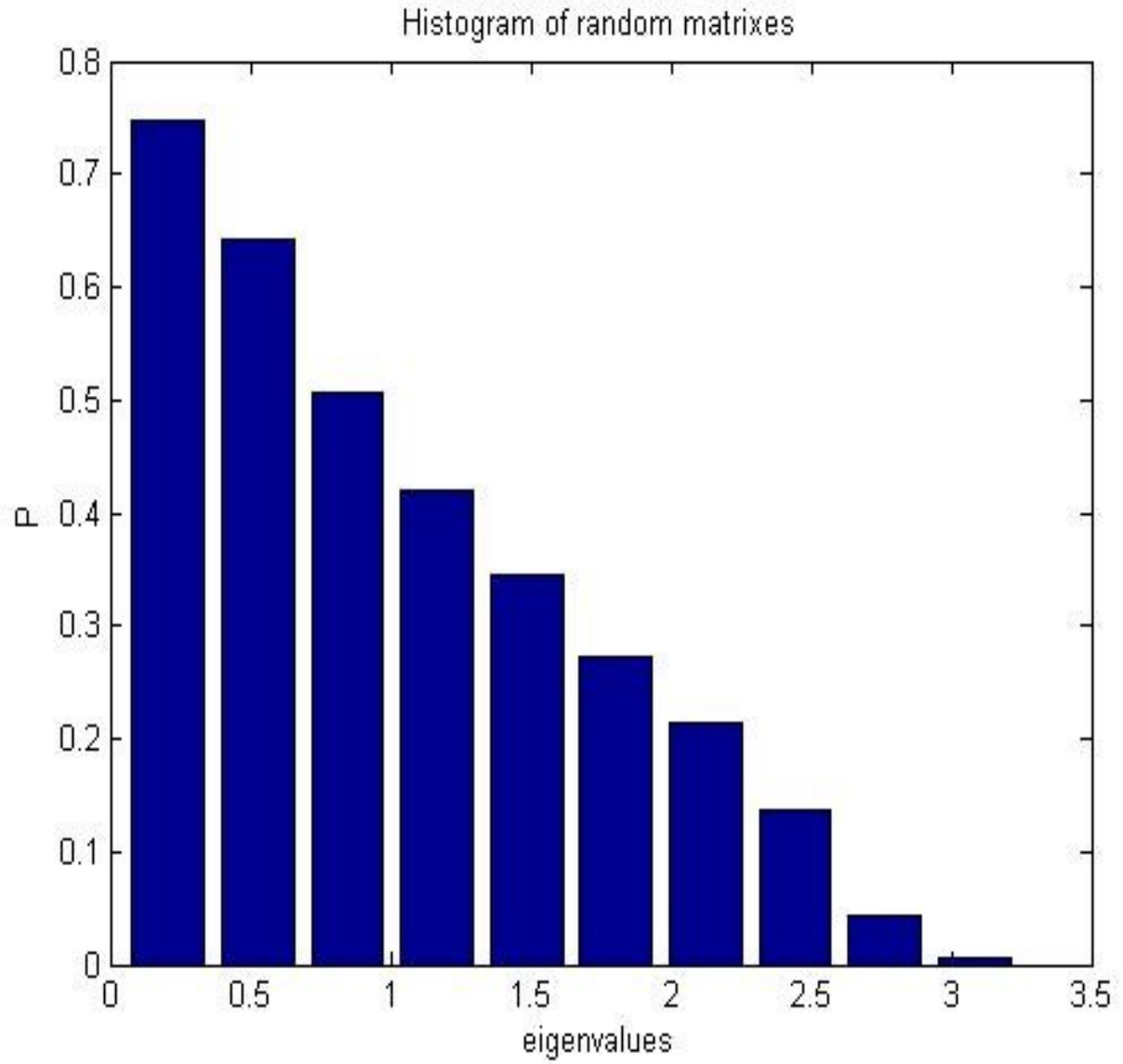
KUSI



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

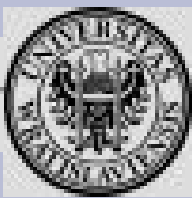
# Random Matrix Theory

Probability of that  
anomaly is less than  
one promil





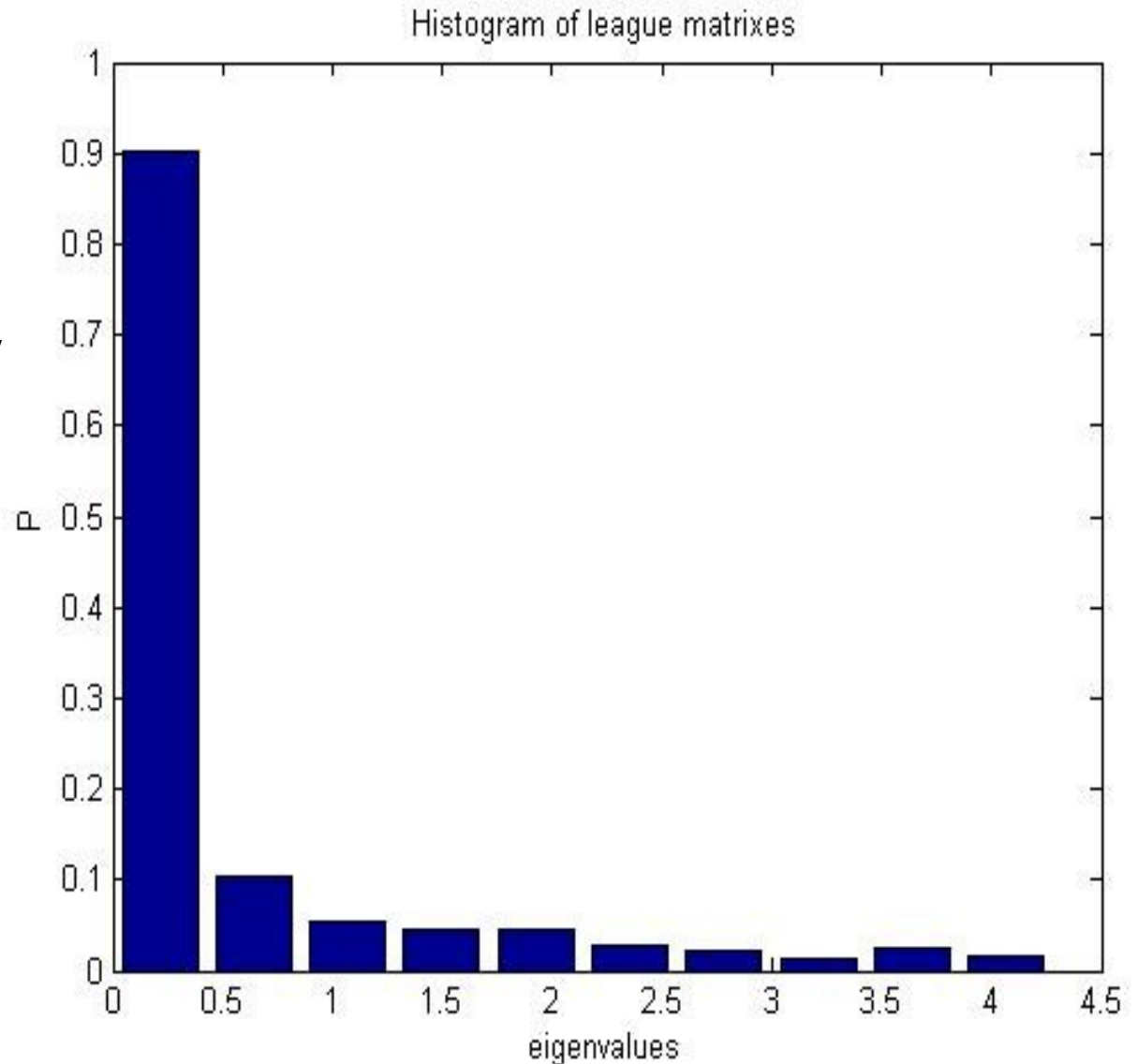
KUSI



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Random Matrix Theory

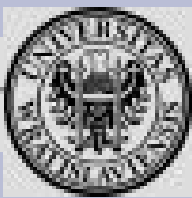
On the other hand histogram of reproduction of smaller (time window 16 matches) league matrixes reduces the possibility of anomaly evidence







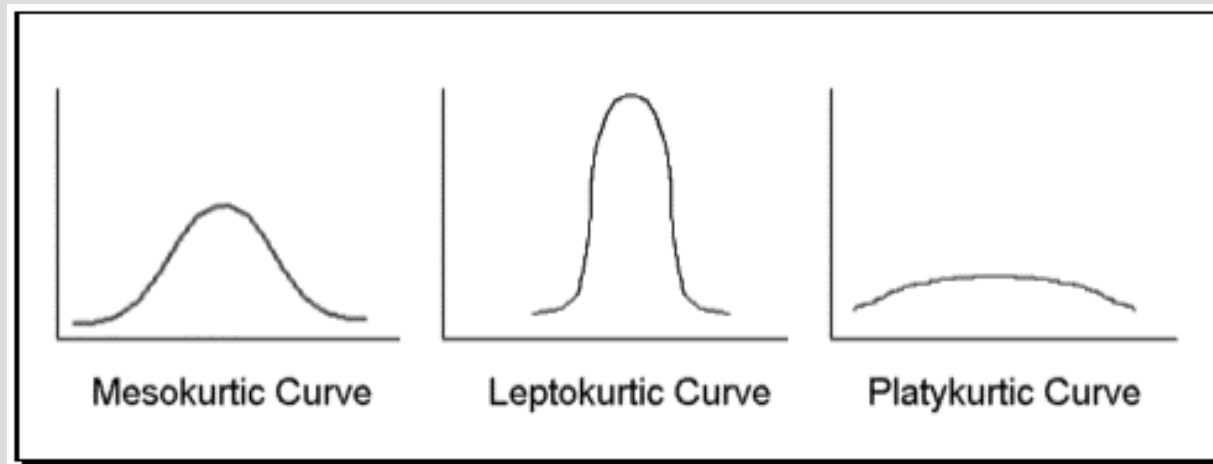
KUSI



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Random Matrix Theory

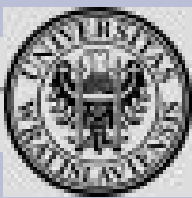
Shape of spectrum density is more leptokurtic if number of rows in matrix is lower



This rule explains that histogram of league matrix is much more leptokurtic than a random one.



KUSI

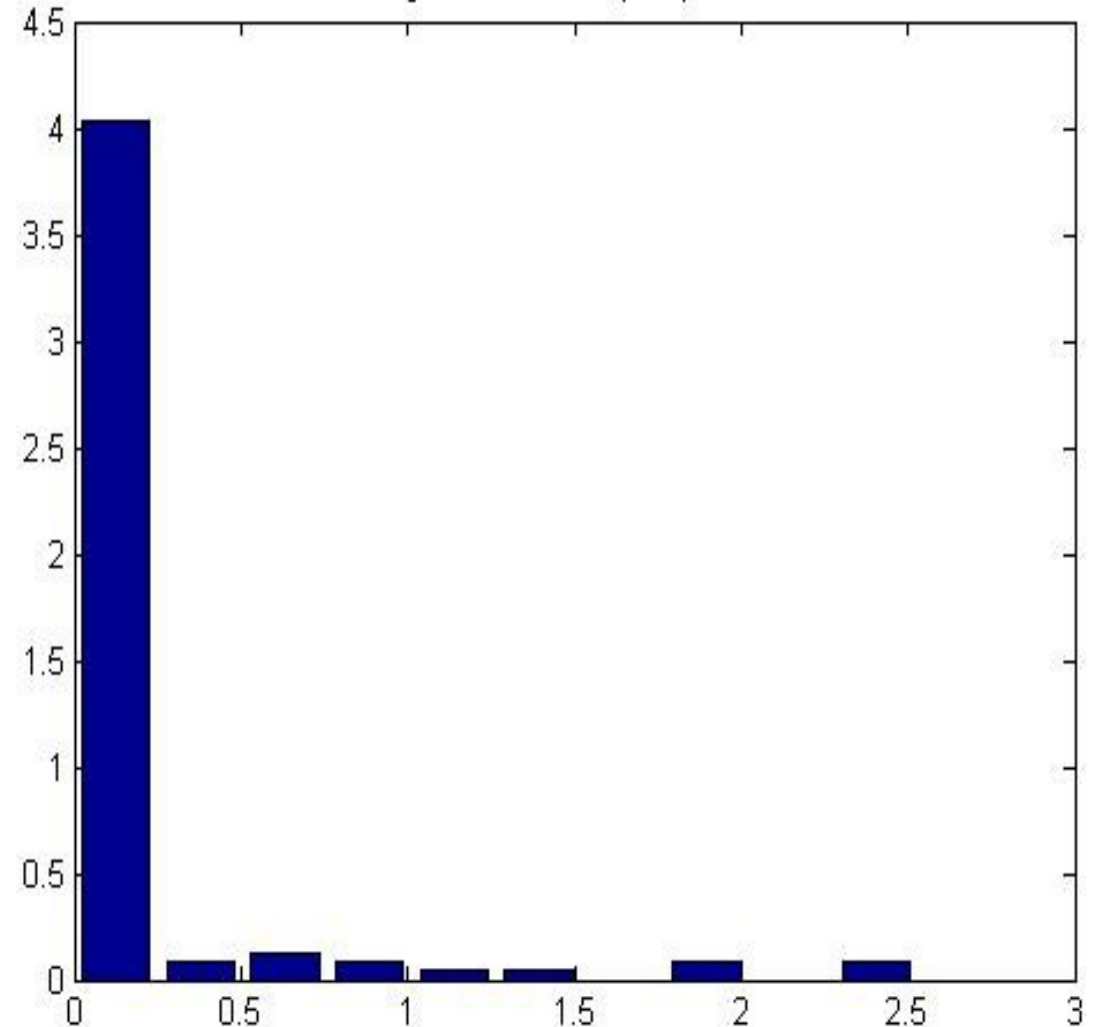


The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Random Matrix Theory

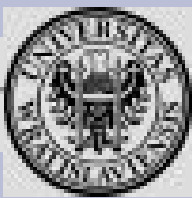
Bootstrap  
replication of  
empirical data

Histogram of bootstrap replications





• KUSI •



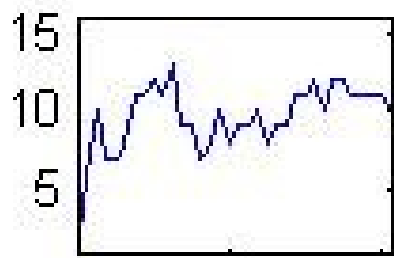
The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Nonlinear measures of dependence

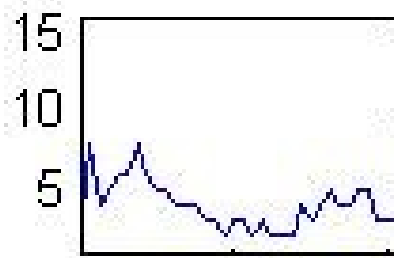
**Ranking analysis:** changing ranks in the table during season

1	Piast Gliwice
2	Cracovia
3	Jagiellonia Białystok
4	GKS Bełchatów
5	Pogoń Szczecin
6	Zagłębie Lubin
7	Tłoki Gorzyce
8	Podbeskidzie Bielsko-Biała
9	Ruch Chorzów
10	Łódzki KS
11	RKS Radomsko
12	Arka Gdynia
13	Aluminium Konin
14	Stasiak Opoczno
15	Polar Wrocław
16	Szczakowianka Jaworzno

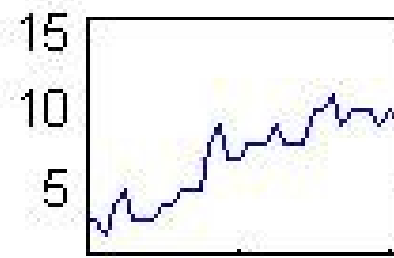
Table of first division  
2003/2004  
after first play



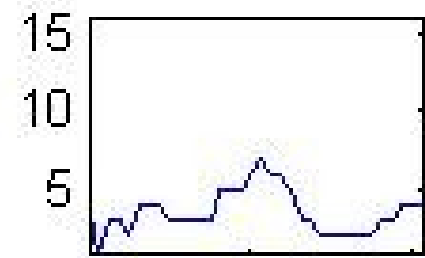
beg half end



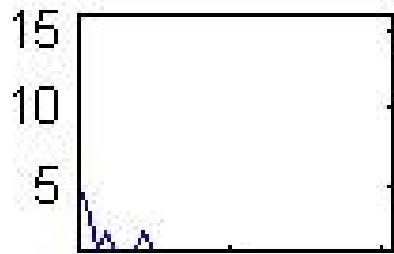
beg half end



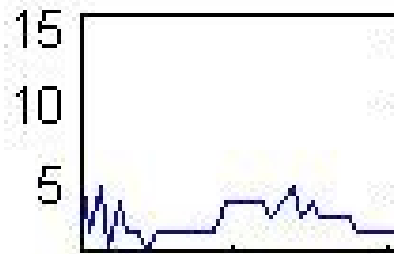
beg half end



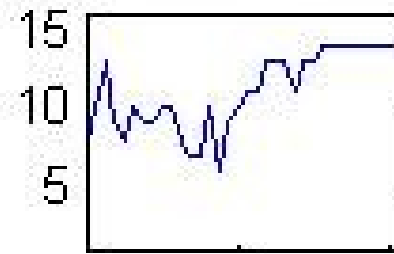
beg half end



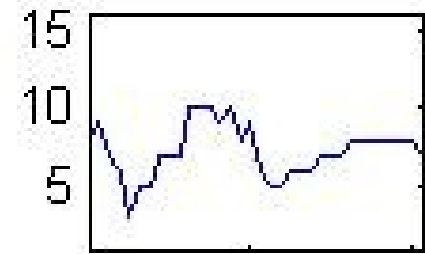
beg half end



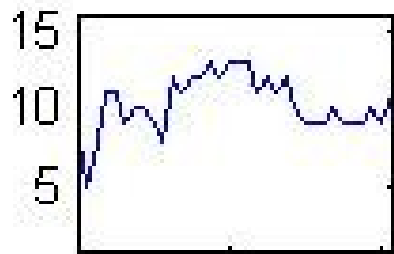
beg half end



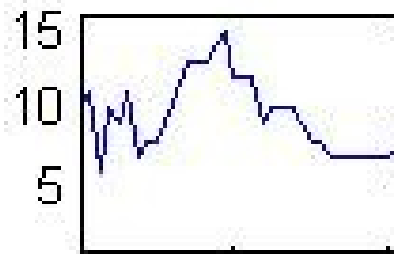
beg half end



beg half end



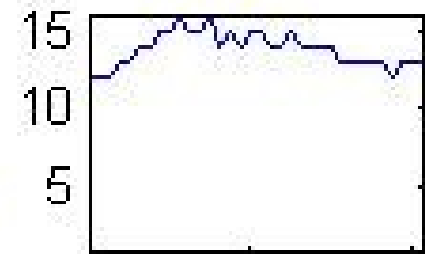
beg half end



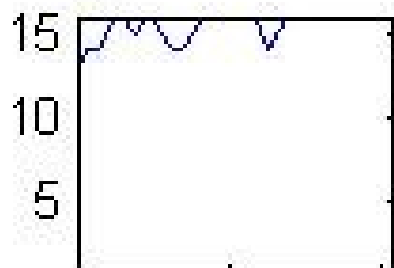
beg half end



beg half end



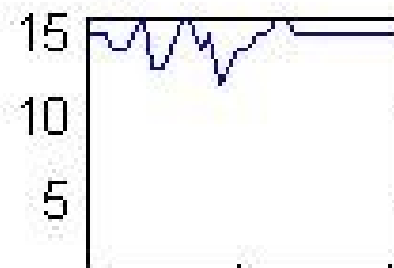
beg half end



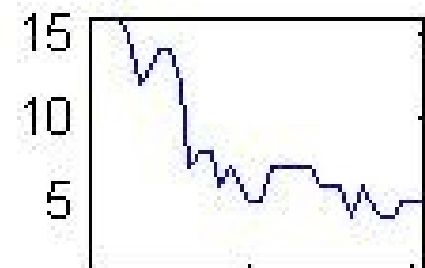
beg half end



beg half end



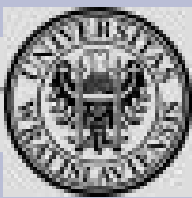
beg half end



beg half end



KUSI

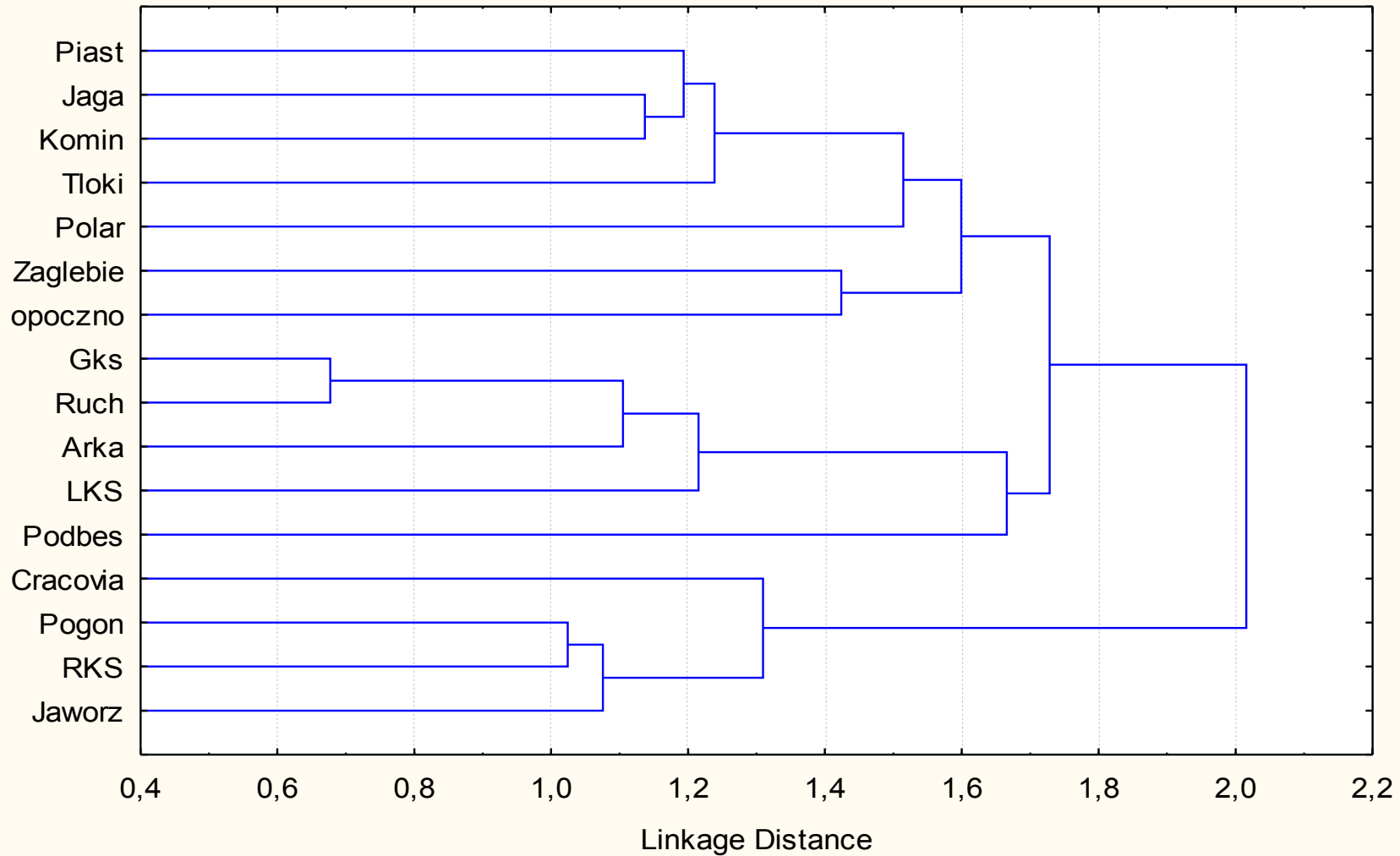


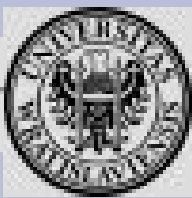
The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Nonlinear measures of dependence

Ranking analyses

Tree Diagram for Variables  
Single Linkage  
Euclidean distances

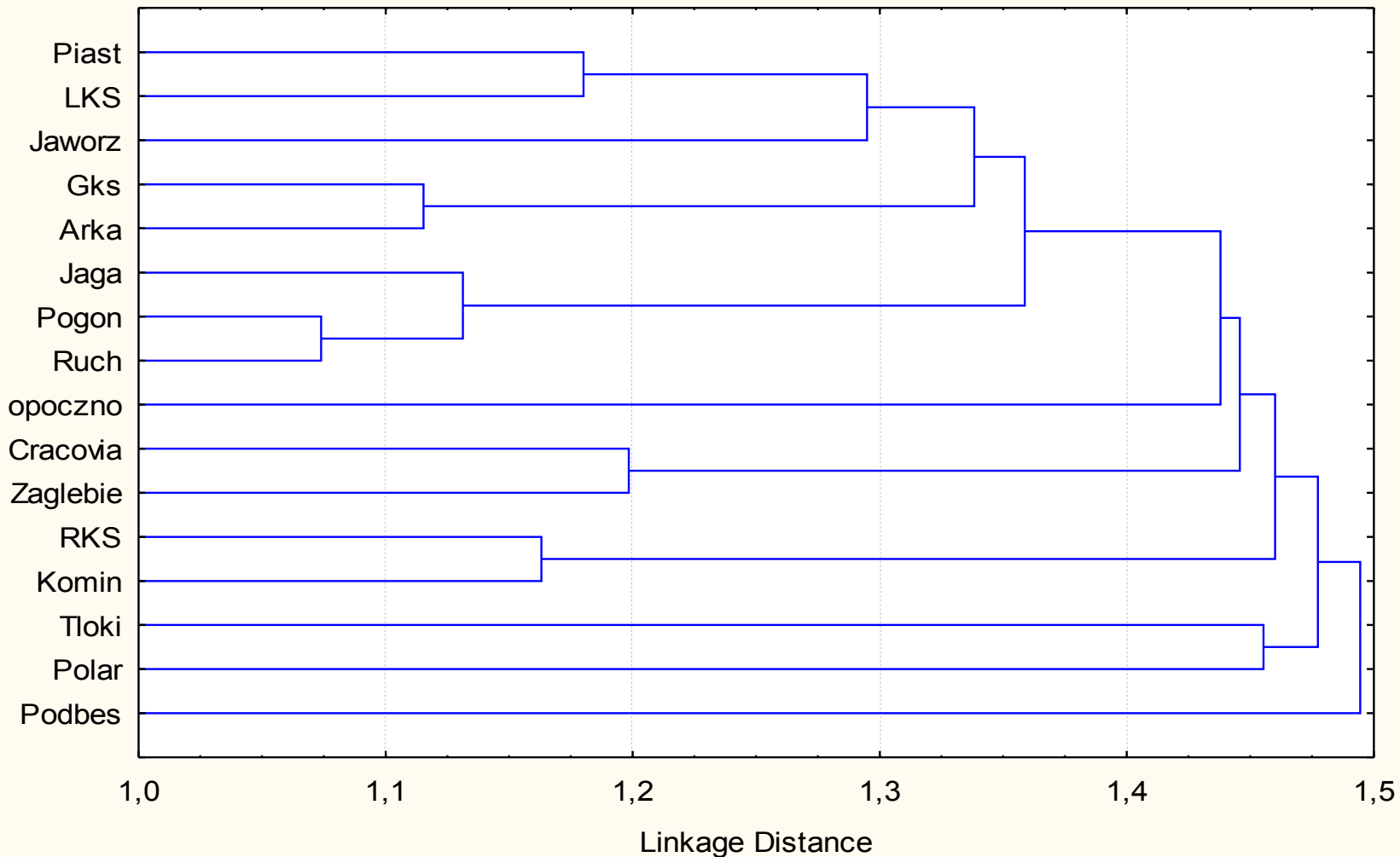




# Nonlinear measures of dependence

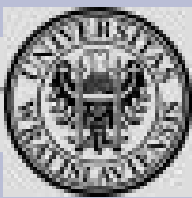
Analyses of results in league  
mesuere 3) - points

Tree Diagram for Variables  
Single Linkage  
Euclidean distances





KUSI

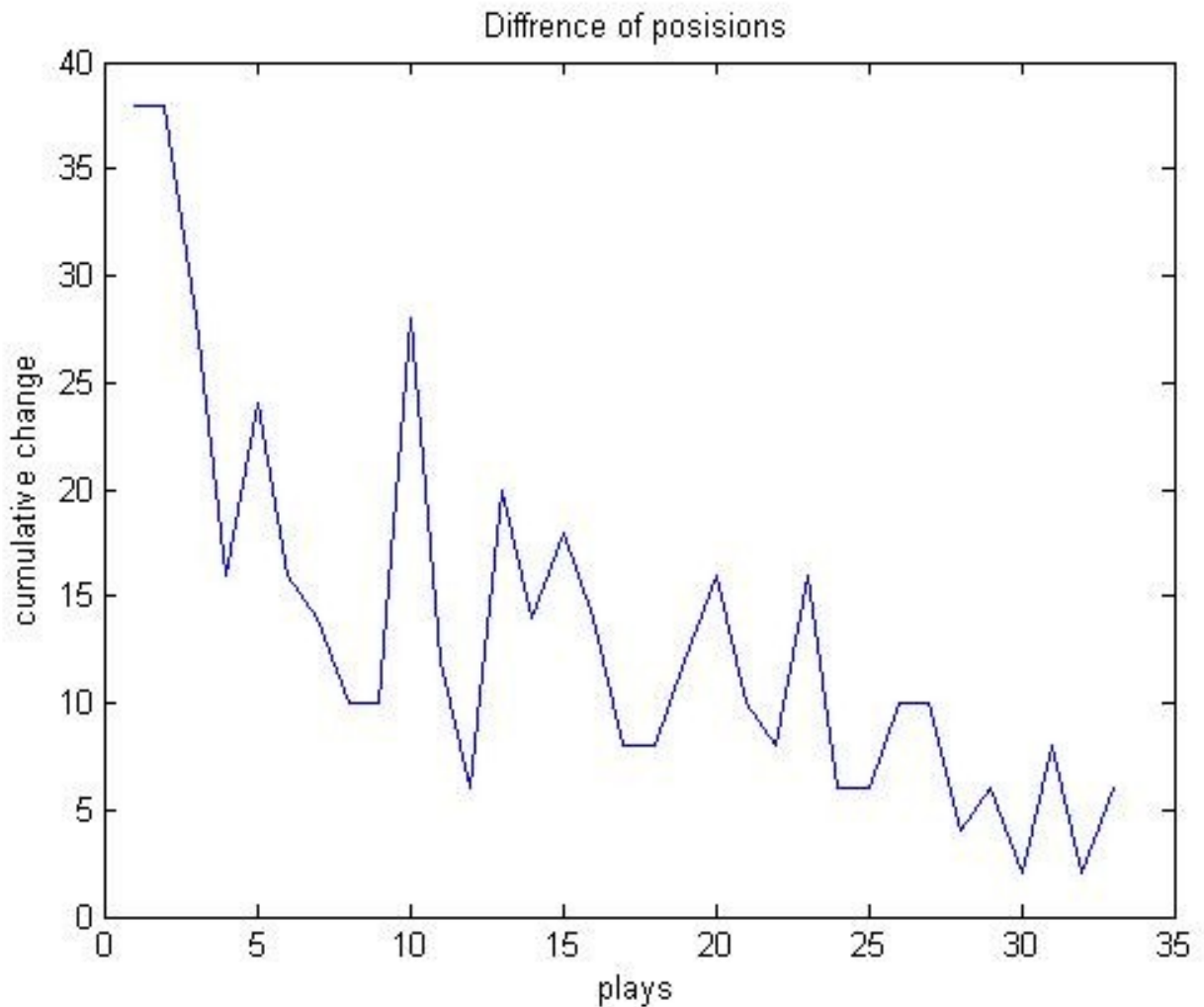


The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Nonlinear measures of dependence

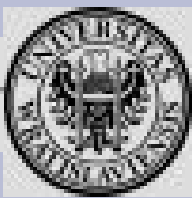
How system goes to equilibrium?

Alternation is decreasing in time with some fluctuations





+ KUSI +



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

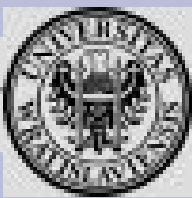
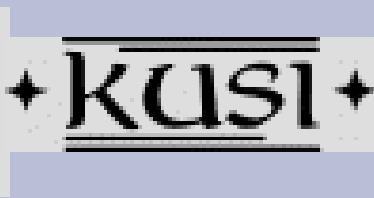
# Nonlinear measures of dependence

**Measure of anomaly:** results which had low probability to occur will be caught by measure construct with difference in table rescaled by special coefficient.

This special scaling coefficient is based on *aprioric* probability of home-team win.

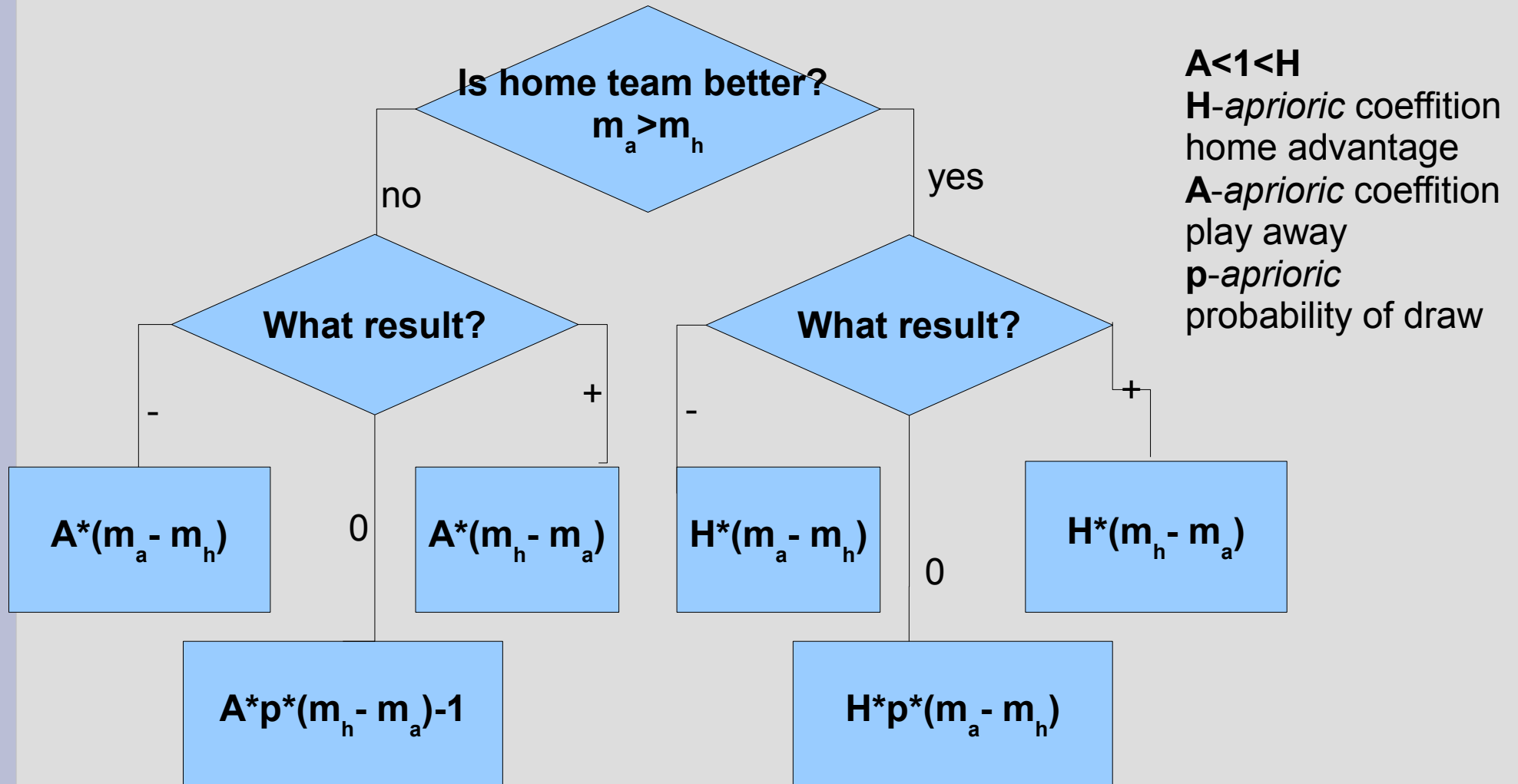
Negative measure means that result was predictable and positive - unpredictable

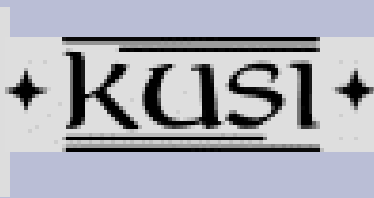




# Nonlinear measures of dependence

Measure construction algorithm:





# Nonlinear measures of dependence

Arka Gdynia

-0,44  
0,02  
-0,12  
0,12  
0,05  
0,1  
-0,48  
0  
0,03  
-0,03  
0,51  
0,13  
-0,03  
-0,02  
0,07  
-0,44  
0,02  
0  
0,12  
-0,08  
-0,24  
-0,36  
0,15  
0  
0,03  
-0,03  
-0,32  
-0,03  
-0,08  
-0,28

Podbeskidzie Bielsko-Biała

0,07  
-0,08  
-0,32  
-0,2  
-0,16  
-0,04  
0,05  
0,08  
-0,12  
-0,16  
0,2  
0,12  
0,05  
0,57  
-0,03  
-0,28  
-0,08  
-0,32  
0,03  
0,02  
-0,04  
-0,24  
0,13  
0,19  
-0,16  
0,32  
0,19  
-0,24  
-0,36  
0,06

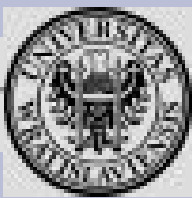
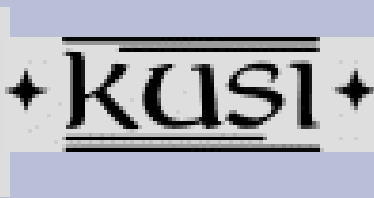
Piast Gliwice

-0,08  
0,04  
-0,16  
-0,12  
-0,03  
-0,32  
-0,28  
-0,12  
0,05  
0,1  
-0,24  
-0,08  
-0,16  
-0,2  
-0,2  
-0,08  
0,06  
0,25  
0  
-0,04  
0,08  
0,44  
0,19  
-0,24  
0,57  
0,05  
-0,08  
-0,16  
0,03  
0,32

Jagiellonia Białystok

0,12  
0,02  
0,28  
-0,03  
0,24  
0,08  
-0,2  
-0,32  
0,07  
0,04  
-0,2  
-0,16  
0,24  
0  
-0,02  
0,19  
0,02  
0,07  
-0,04  
0,05  
0,13  
-0,2  
-0,32  
0,44  
-0,03  
0,03  
0,25  
-0,24  
0  
-0,02

Chosen teams  
and theirs  
measures in  
time sequence



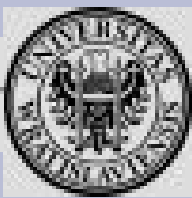
# Nonlinear measures of dependence

## Comparable table - Bundesliga 2007/2008

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1: VfB Stuttgart	0	0,06	-0,11	-0,17	-0,22	0,09	-0,33	-0,39	0,44	0,16	-0,56	-0,61	-0,67	-0,72	0,25	-0,83	-0,89	0,31
2: FC Bayern München	-0,04	0	0,02	0,11	0,17	0,07	0,09	-0,33	0,13	0,14	0,16	-0,56	-0,61	0,22	0,24	-0,78	0,27	0,89
3: Bayer 04 Leverkusen	-0,07	-0,04	0	-0,06	-0,11	-0,17	0,07	0,09	0,33	-0,39	-0,44	-0,5	0,18	-0,61	0,67	-0,72	0,78	0,83
4: 1. FC Nürnberg	0,11	0,07	0,04	0	-0,06	-0,11	-0,17	0,07	-0,28	0,33	-0,39	-0,44	0,16	0,18	-0,61	0,67	-0,72	-0,78
5: VfL Bochum	0	-0,11	-0,07	0,04	0	-0,06	-0,11	-0,17	-0,22	-0,28	-0,33	-0,39	0,14	-0,5	-0,56	-0,61	-0,67	-0,72
6: Borussia Dortmund	0,01	-0,14	-0,11	-0,07	-0,04	0	0,02	-0,11	0,17	0,07	-0,28	0,11	0,13	-0,44	0,16	-0,56	-0,61	0,67
7: Hannover 96	-0,21	0,01	-0,14	-0,01	-0,07	0,04	0	0,06	0,11	-0,17	-0,22	0,28	0,11	0,13	0,44	-0,5	-0,56	-0,61
8: Eintracht Frankfurt	0,25	0,21	0,18	0	-0,01	0,07	0,04	0	0,06	0,04	0,05	0,22	0,09	0,11	0,39	0,44	0,5	-0,56
9: VfL Wolfsburg	-0,28	0,04	0,02	-0,18	0	0,11	0,07	-0,03	0	0,06	-0,11	0,17	0,07	-0,28	0,33	-0,39	0,14	0,5
10: FC Schalke 04	0,32	-0,28	0,25	0,21	0,01	0,14	-0,01	-0,02	0,04	0	-0,06	-0,11	0,05	-0,22	0,28	0,11	0,13	0,14
11: Werder Bremen	-0,36	0,06	0,28	0,04	-0,21	0,18	0,14	0,11	-0,07	-0,03	0	0,06	0,11	-0,17	0,07	0,28	-0,33	0,13
12: Hamburger SV	0,39	0,07	-0,32	-0,28	-0,25	0,21	0,18	0	-0,11	-0,02	-0,04	0	0,06	0,11	0,05	0,22	-0,28	-0,33
13: Hertha BSC Berlin	0,43	0,08	0,36	-0,32	-0,28	0,25	0,02	0,01	0,14	-0,01	-0,07	0,04	0	0,06	0,11	0,05	0,22	0,28
14: Arminia Bielefeld	-0,46	0,43	-0,39	0,07	0,32	0,28	-0,25	0,21	0,18	0,14	-0,11	0,07	0,04	0	0,06	0,11	0,05	0,07
15: FC Energie Cottbus	0,5	-0,46	0,43	-0,39	-0,36	0,32	0,28	0,25	0,21	0,18	0	0,11	-0,07	0,04	0	0,06	-0,11	-0,17
16: Karlsruher SC	0,53	0,5	-0,46	-0,43	0,39	0,36	0,06	0,28	0,04	0,02	-0,18	0	0,11	-0,07	-0,04	0	0,02	0,04
17: FC Hansa Rostock	0,57	0,13	0,12	-0,46	-0,43	-0,39	0,36	0,06	0,28	0,25	0,21	0,18	-0,14	-0,01	0,07	0,04	0	0,06
18: MSV Duisburg	-0,6	0,57	-0,53	-0,5	0,46	0,43	-0,39	-0,36	0,32	-0,28	-0,25	0,02	0,18	-0,14	-0,11	0,07	0,04	0



• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

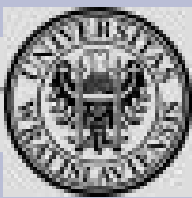
# Nonlinear measures of dependence

**3<sup>rd</sup> quantile of empirical distribution  
of our measure:**

- 0,14- German league
- 0,12- Polish league



• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Nonlinear measures of dependence

**3<sup>rd</sup> quantile of empirical distribution of our measure:**

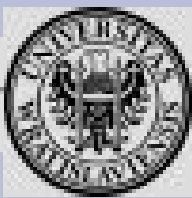
- 0,14- German league
- 0,12- Polish league

Polish league is less predictable than German!

*Maybe we should set alternative hypothesis of corruption?*



KUSI



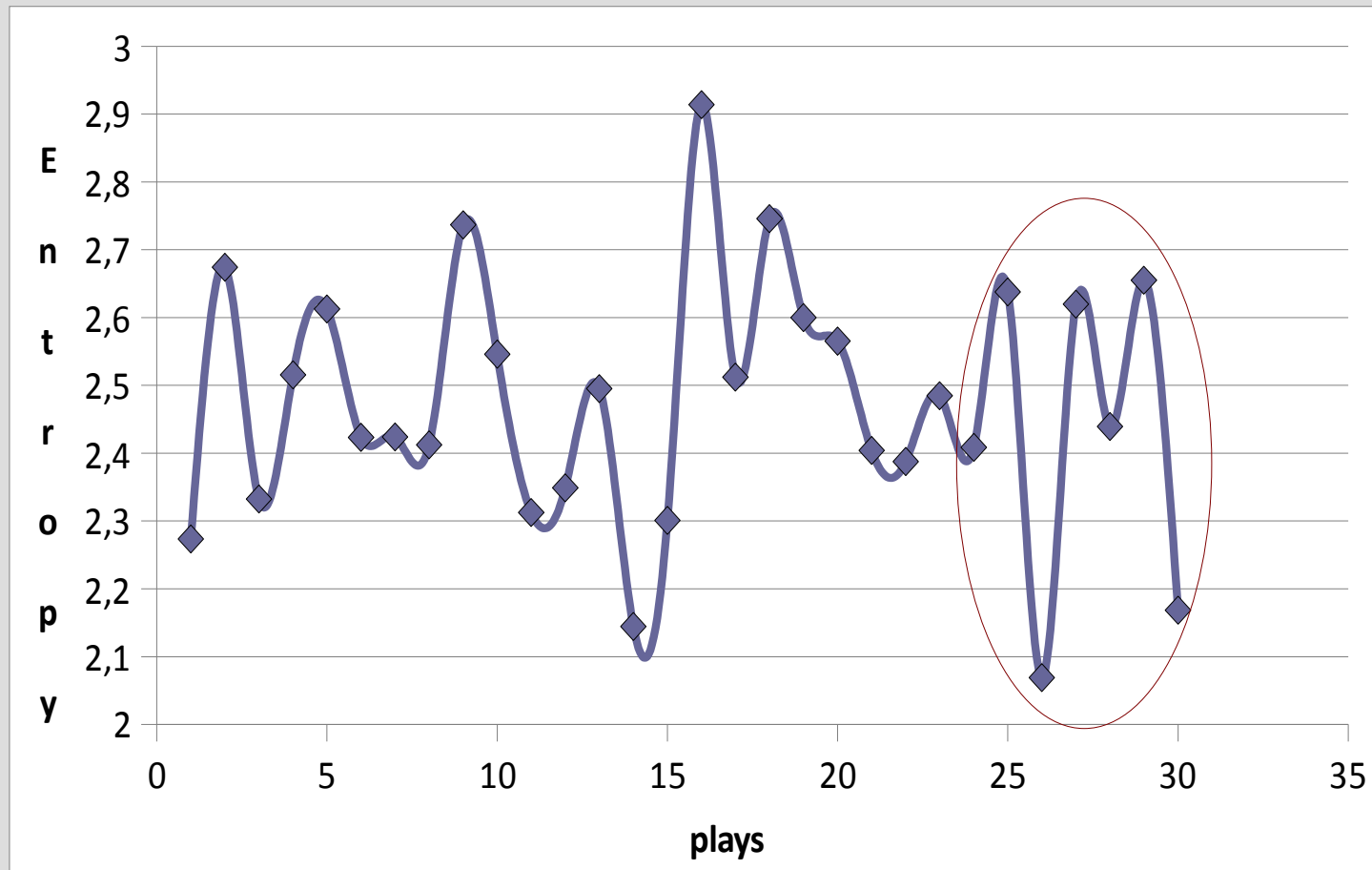
The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Nonlinear measures of dependence

Let consider: our mesurment will play role of probabilities after rescaling to (0,1)

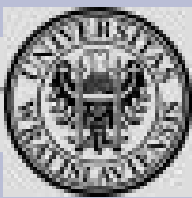
Shanon  
quasi-Entropy:

$$qS = - \sum p_i \ln(p_i)$$





KUSI

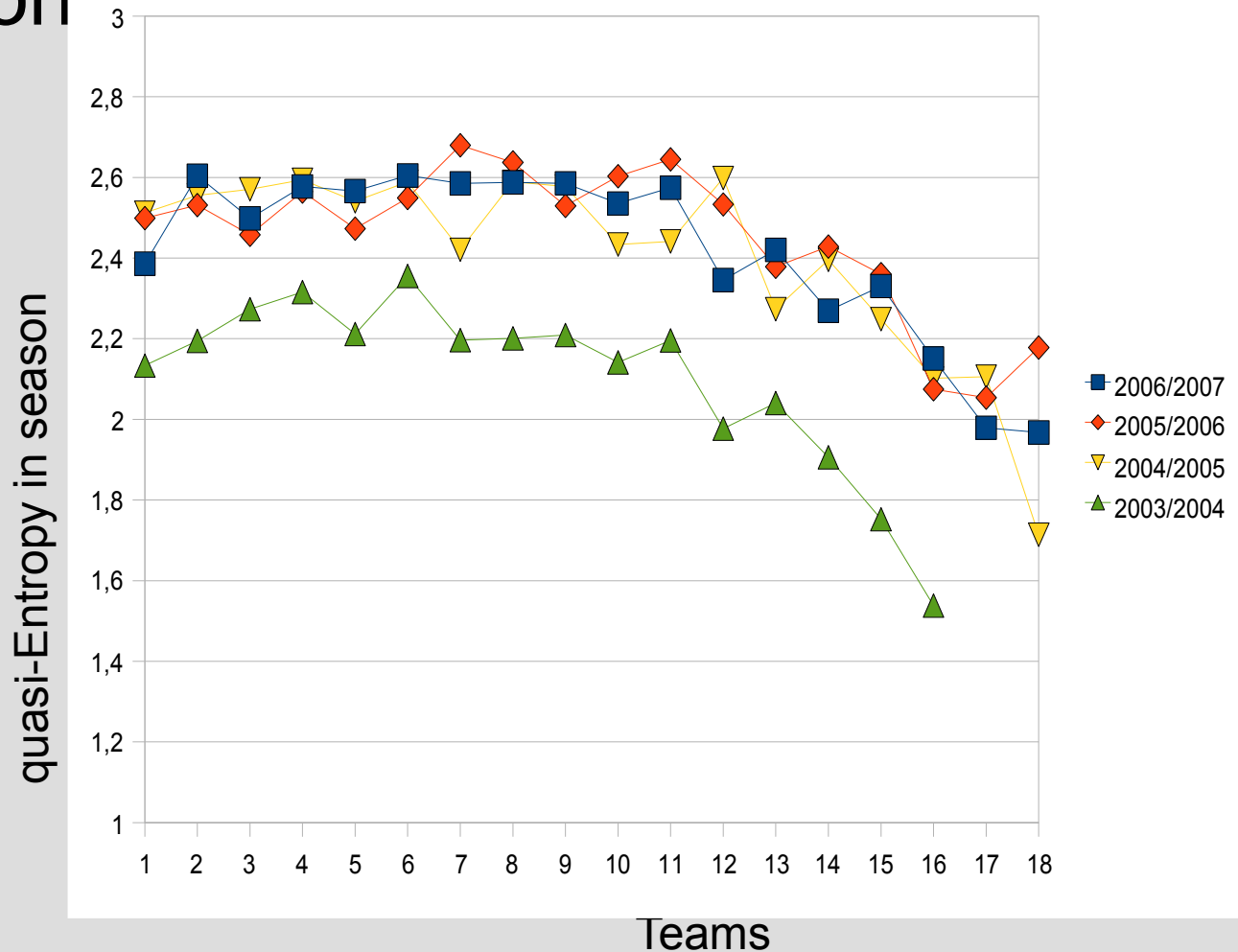


The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# Nonlinear measures of dependence

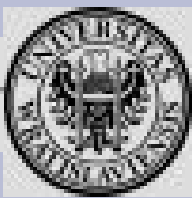
Entropy for teams  
in polish first division

Teams from end  
of table are more  
unpredictable





• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

# The End

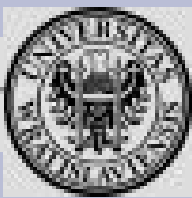
Thanks to my supervisor – dr Janusz Miśkiewicz

I really regret, that I didn't proof anything about corruption





• KUSI •



The UNESCO Chair of Interdisciplinary Studies  
University of Wrocław

## The End

Thanks to my supervisor – dr Janusz Miśkiewicz

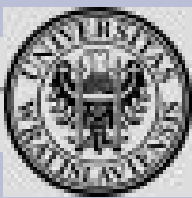
I really regret, that I didn't proof anything about corruption

but I'm still working on this problem...

Data:

<http://www.bundesliga-statistik.de/>

<http://www.90minut.pl/>



# The End

## Appendix: Game Theory

$\frac{1}{2}$	D	A
D	$(W1-L, W2+G-C-m)^*$	$W1+G-C-m, W2-L+C-m$
A	$W1-L+C-m, W2+G-C-m$	$W1, W2$

$W2 = a_{21}$  +/- home-field advantage

$W1 = a_{12}$  +/- home-field advantage

G- profit in points

L- lost in points

C- inducement

m- moral cost