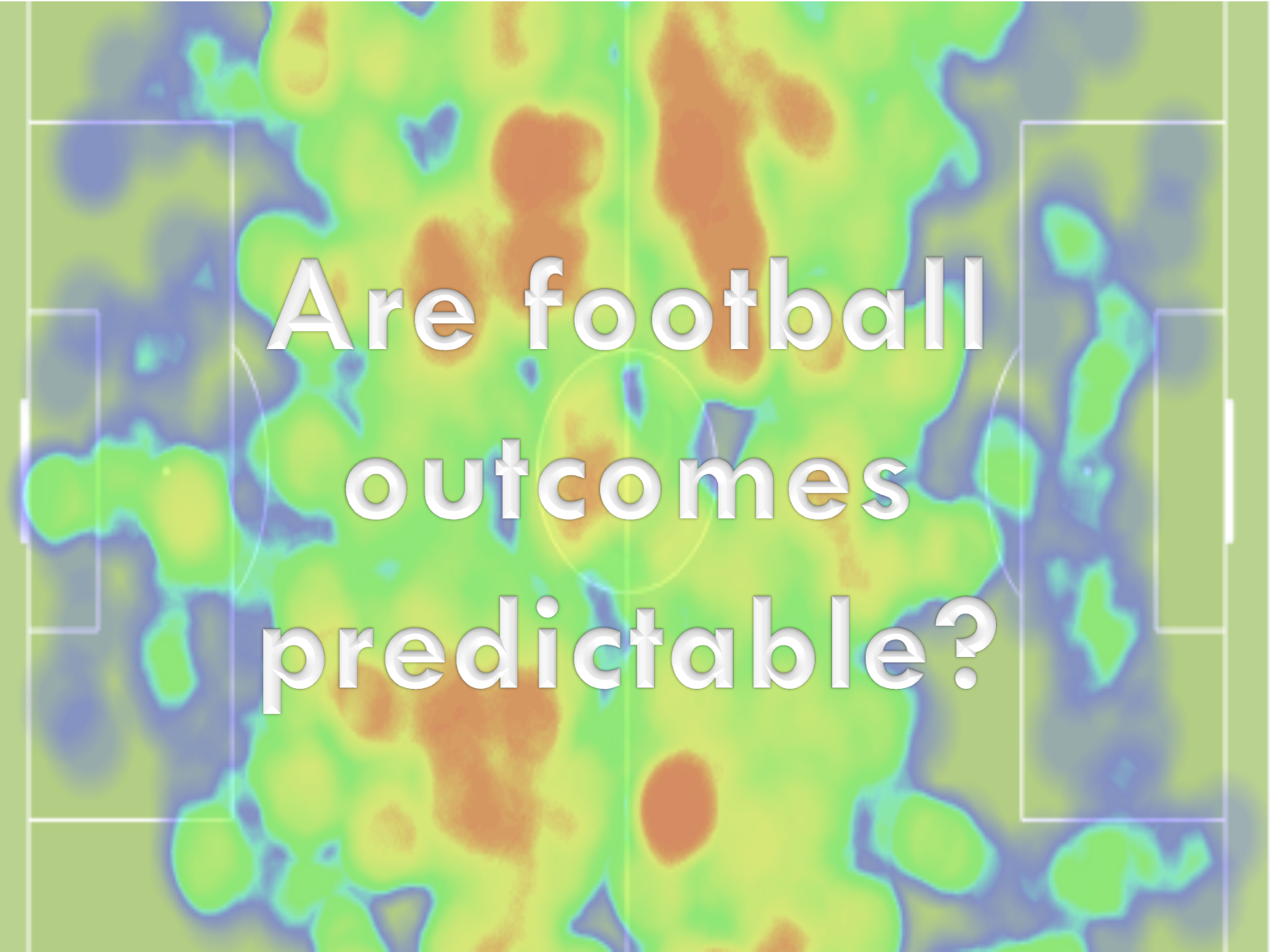
A NETWORK-BASED APPROACH TO EVALUATE THE PERFORMANCE OF FOOTBALL TEAMS

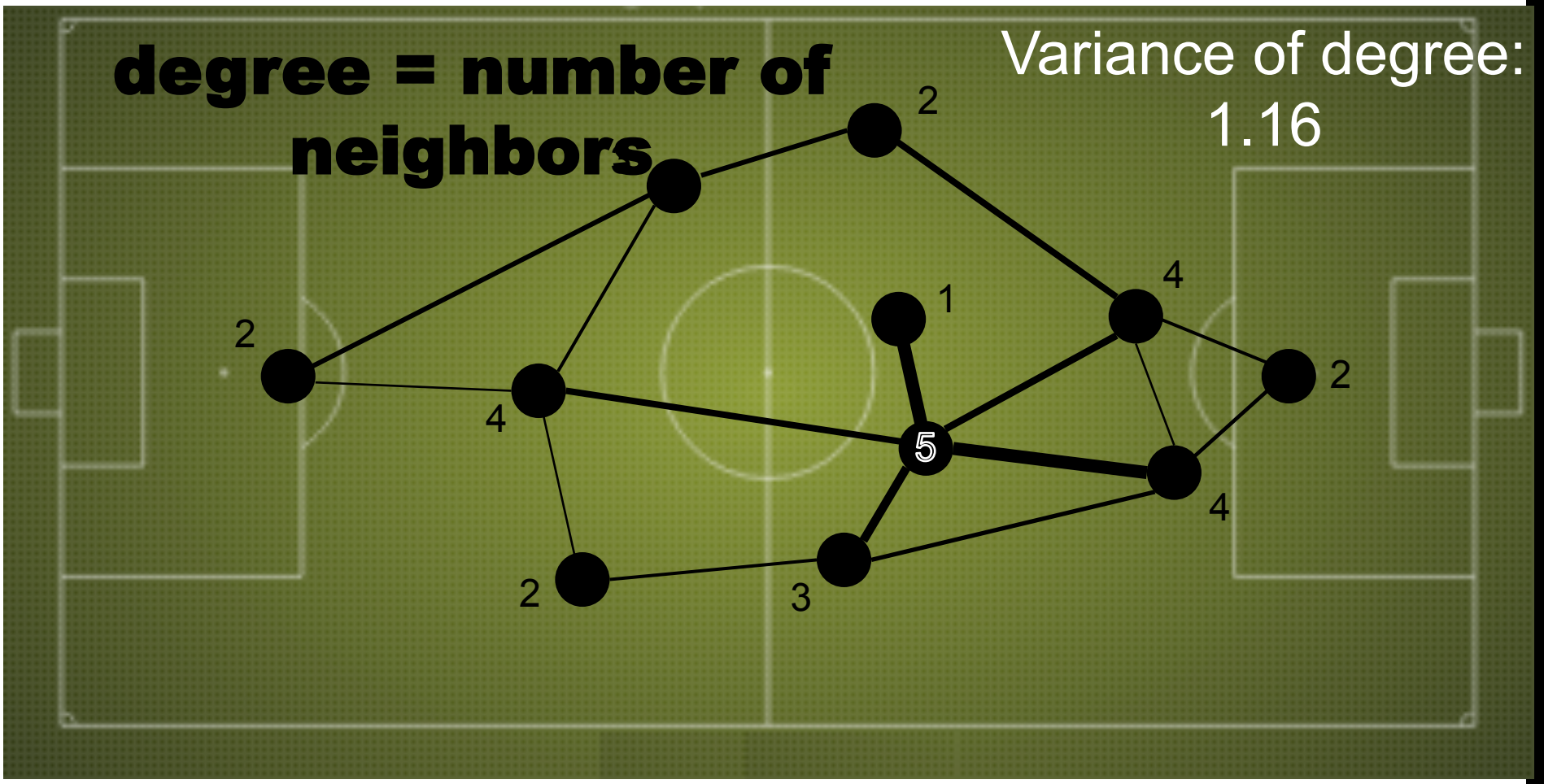
P. Cintia, L. Pappalardo, S. Rinzivillo, D. Pedreschi, F. Giannotti
{name.surname}@isti.cnr.it
http://kdd.isti.cnr.it


Are football
outcomes
predictable?

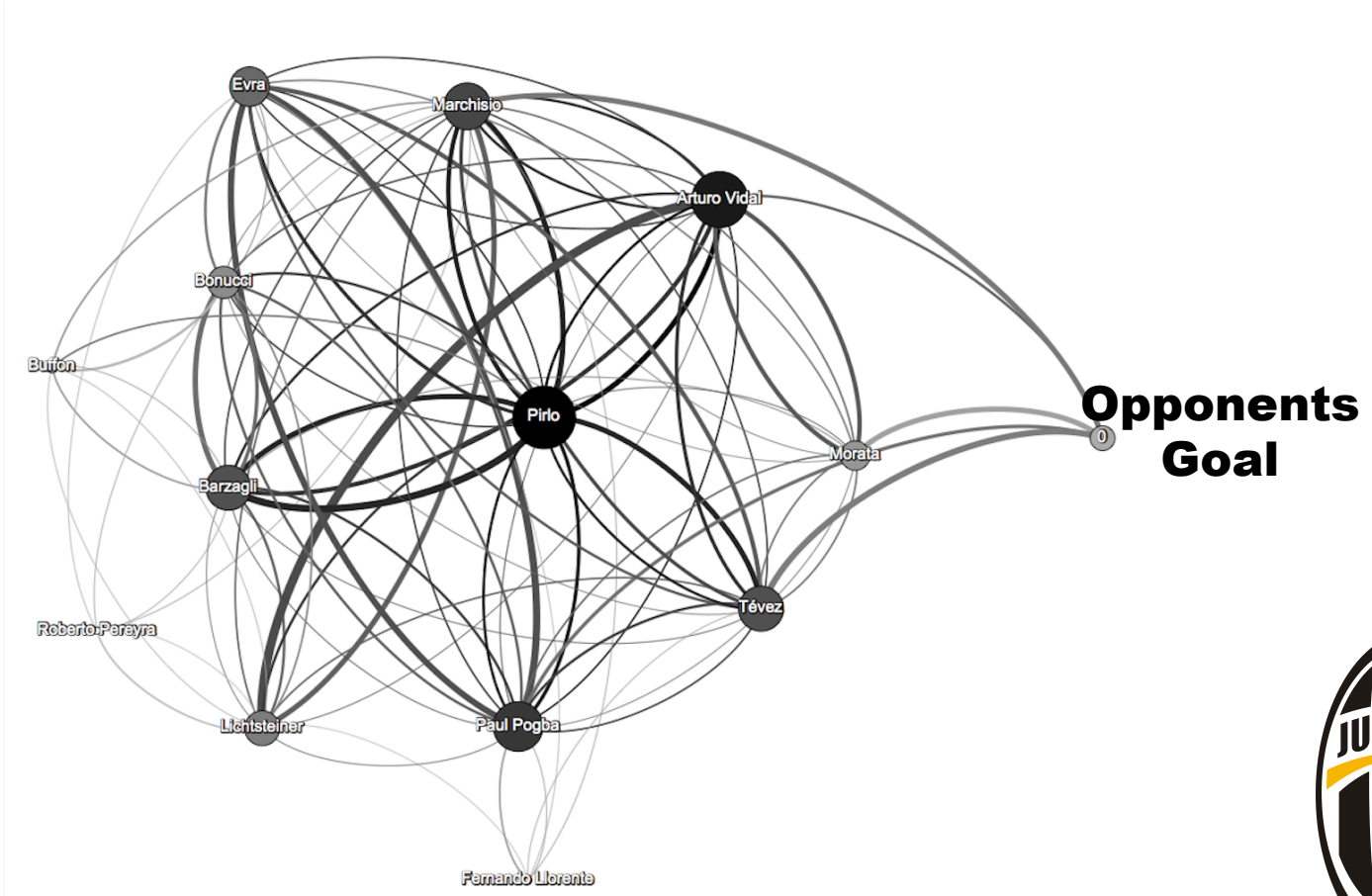
Data from each single match

```
...
<tackle,15.4,41.1,112>
<pass,25.0,67.1,113>
<pass,65.0,87.1,115>
<assist,82.1,35.8,120>
<goal attempt,82.1,35.8,121>
.....
```

THE PASSES NETWORK AMONG PLAYERS

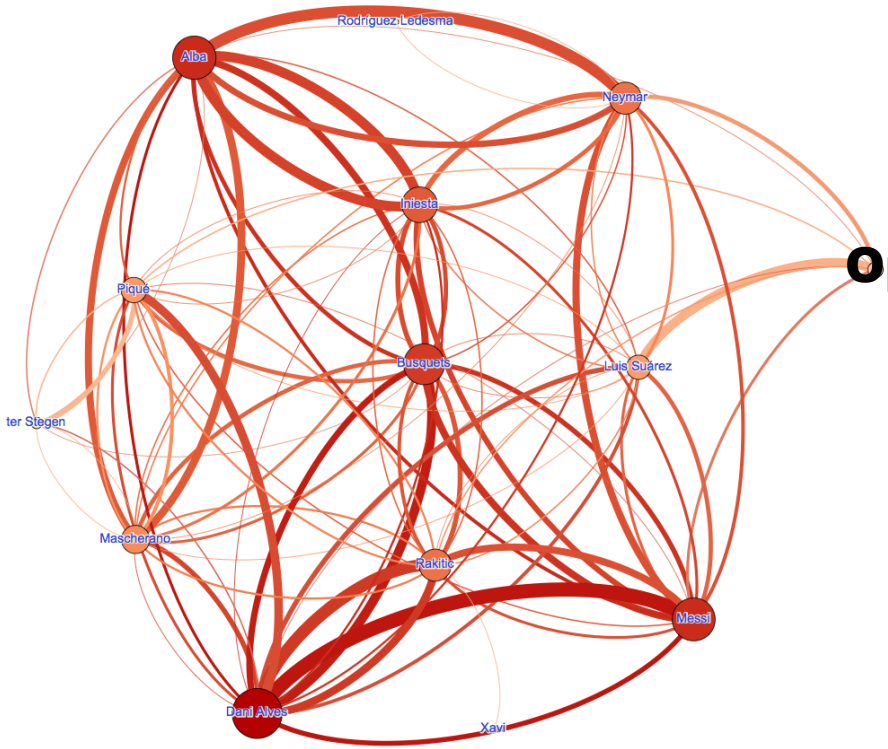


FOOTBALL AS A NETWORK



Juventus passes network from last champions league game

FOOTBALL AS A NETWORK



Opponents Goal

Barcelona passes network from last champions league game

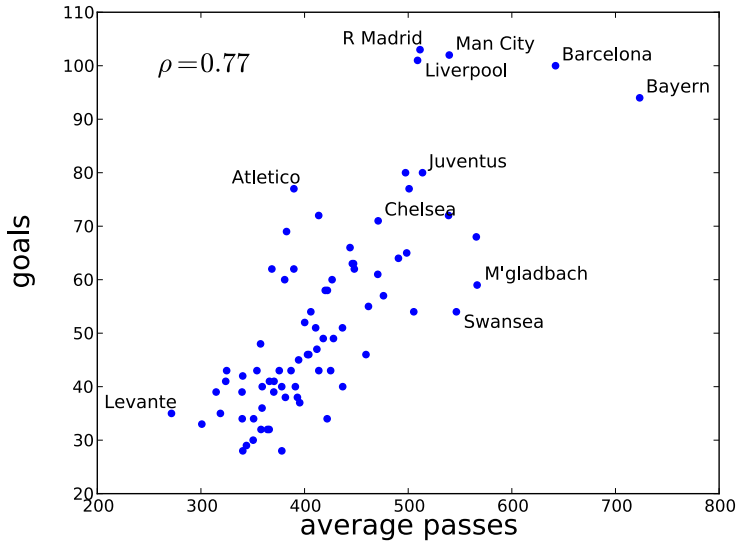
NETWORK ANALYSIS FOR PERFOMANCE EVALUATION

- Networks characteristics are a proxy for performance evaluation and prediction
- We use only passing networks to outperform the results of standard predictors

measure	description
w	total passing volume
μ_p	mean players' passing volume
σ_p	variance of players' passing volume
μ_z	mean zones' passing volume
σ_z	variance of zones' passing volume
H	combination of above measures

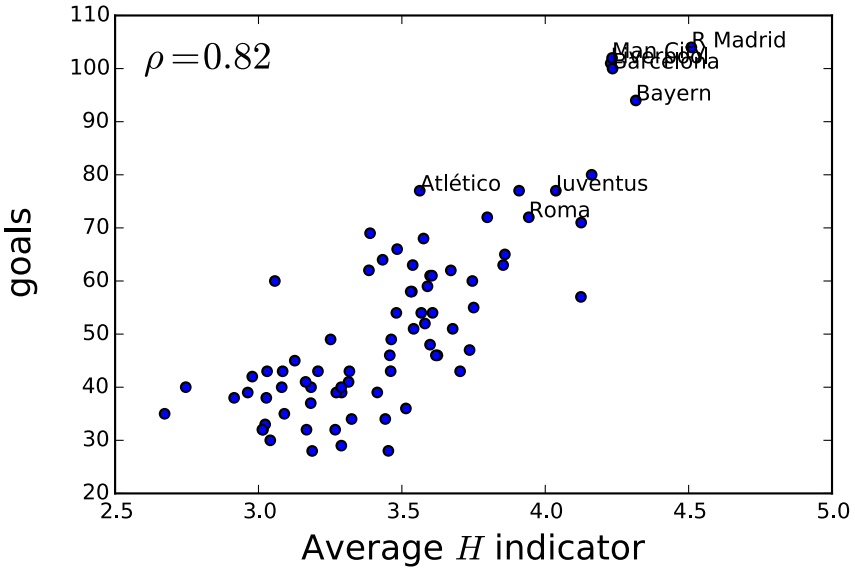
Measures involved in our model: we combine different passing indexes into one single indicator (H)

EVALUATING THE EVALUATOR



Using the average passes per match, the correlation with goals is 0.77…

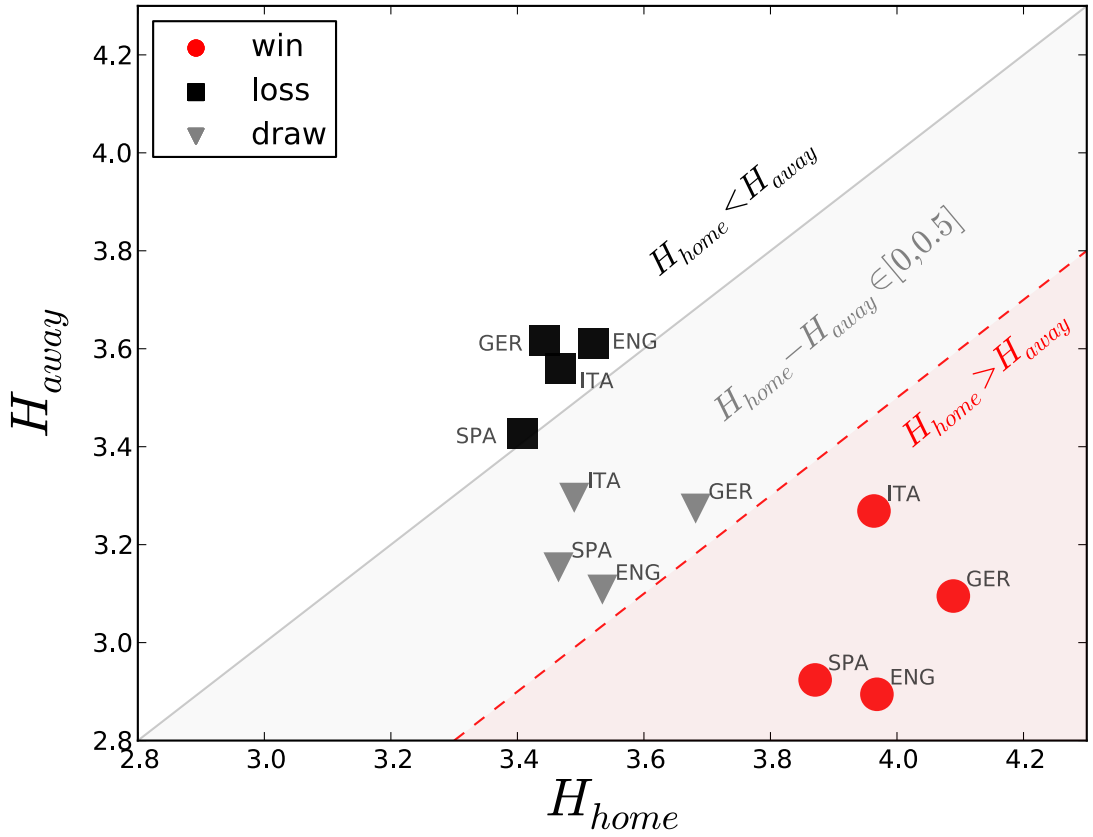
…while the H indicator has a correlation with goals equal to 0.82



H INDICATOR IN EUROPEAN LEAGUES

	team	mean *H*	league
1	Real Madrid	4.51	SPA
2	Bayern München	4.31	GER
3	Barcelona	4.23	SPA
4	Manchester City	4.23	ENG
5	Liverpool	4.22	ENG
6	Borussia Dortmund	4.16	GER
7	Chelsea	4.12	ENG
8	Milan	4.12	ITA
9	Juventus	4.03	ITA
10	Roma	3.94	ITA

ASSESSING TEAM PERFORMANCES



For each game we consider the H indicator of both teams and we cluster this points according to the real outcome. Centroids of such clusters are confirming the goodness of our approach.

FOOTBALL GAMES PREDICTION

- We train several prediction model with a dataset composed by H indicator of teams and we try to predict games outcome
- We used the best result from three dummy classifiers (random, class distribution, most frequent label) as baseline
- We have cross-validated the results of each classifier

classifier	Germany	England	Spain	Italy
KNearestNeighbor	**0.60**	0.55	0.51	0.52
Logistic Regression	0.53	0.57	0.52	0.53
Decision Tree	0.54	0.56	0.50	0.53
SVM	0.53	0.57	0.52	0.53
Naive Bayes	0.50	0.56	0.49	0.50
Random Forest	0.57	**0.58**	**0.53**	**0.55**
baseline	0.45	0.45	0.45	0.45

Results of our predictions for the main football leagues

THANKS!

Follow us on Twitter: @bigdatatales

Luca Pappalardo @siffolone

Paolo Cintia @mesosbrodleto

Salvatore Rinzivillo @rinziv

simulated ranking		real ranking	
Bayern	91	**Bayern**	90
Leverkusen	72	Dortmund	71
Dortmund	68	Schalke	64
Wolfsburg	59	Leverkusen	61
Augsburg	58	Wolfsburg	60
Hoffenheim	49	Mönchengladbach	55
Hertha	49	Mainz	53
Mainz	48	Augsburg	52
Schalke	47	Hoffenheim	44
Frankfurt	46	Hannover	42
Mönchengladbach	42	Hertha	41
Hannover	41	Werder	39
Hamburg	38	Freiburg	36
Stuttgart	35	Frankfurt	36
Freiburg	31	Stuttgart	32
Werder	24	Hamburg	27
Braunschweig	22	Nürnberg	26
Nürnberg	17	Braunschweig	25

“E’ la dura legge del gol

fai un gran bel gioco però

se non hai difesa

gli altri segnano…

…e poi vincono.”

Max Pezzali, 1998

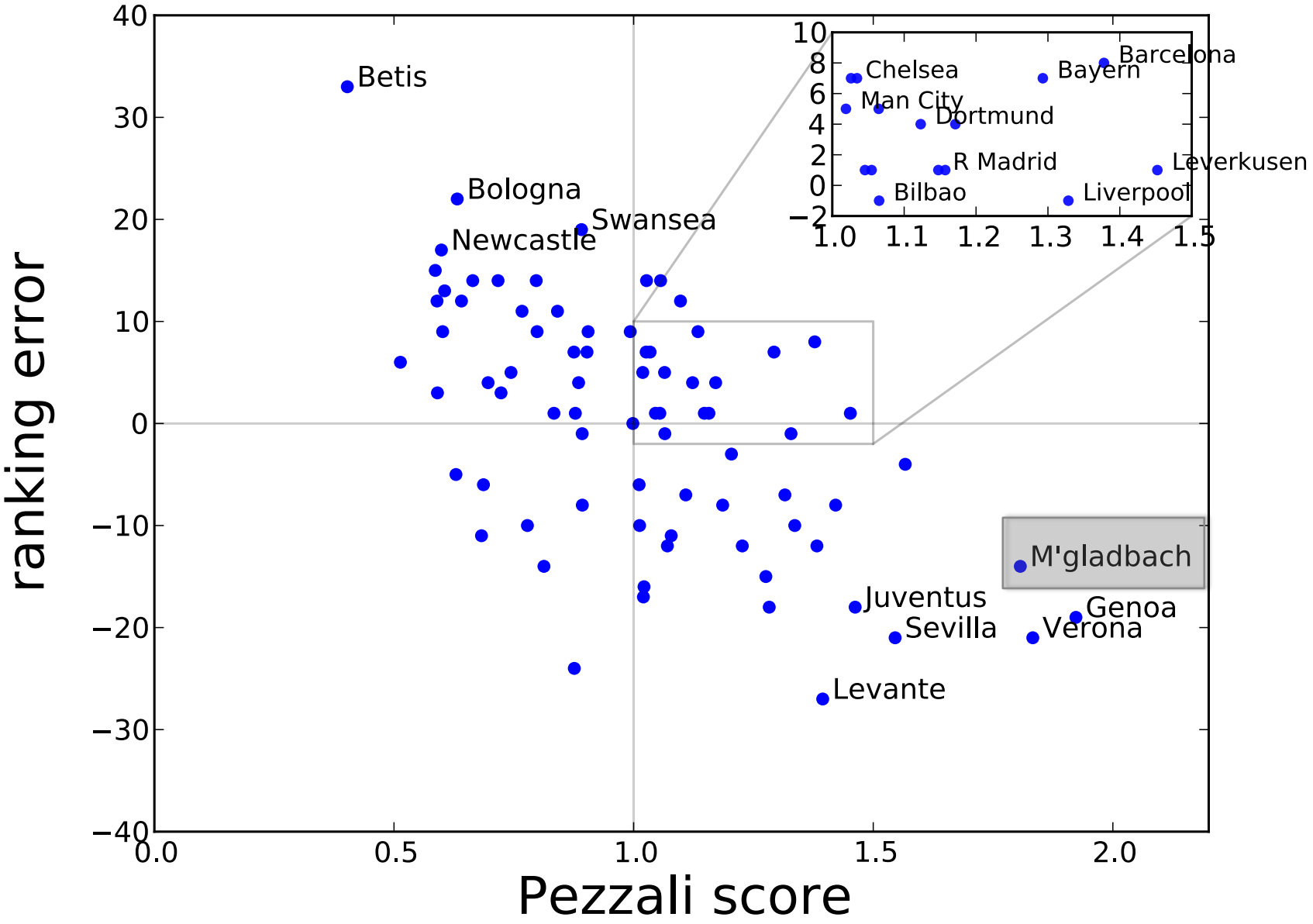
The harsh (mathematic) law of the goals

$$Pezzali\,score(A) = \frac{\sum_A g_A}{\sum_A t_A} * \frac{\sum_B t_B}{\sum_B g_B}$$

A,B= team A, team B

g: goals
t: attempts

Avg Inter: 0.4
Avg Juventus: 1.5


ranking error
Pezzali score
Betis
Bologna
Swansea
Newcastle
M'gladbach
Juventus
Genoa
Sevilla
Verona
Levante
Chelsea
Barcelona
Bayern
Man City
Dortmund
R Madrid
Leverkusen
Bilbao
Liverpool