A Probabilistic Model for Multi-Contestant Races

Panos Pardalos

Distinguished Professor

CAO, Dept. of Industrial and Systems Engineering, University of Florida

and

LATNA, National Research University Higher School of Economics

http://www.ise.ufl.edu/pardalos

ATINER Conference

May 2016

Athens, Greece

Summarizing and Visualizing of Sports Data

- Sports competitions generate massive amounts of data
- This data must be efficiently analyzed and visualized

Relationships in sport networks

- In sports networks there are many relationships we can consider:
- **athlete - athlete** (cooperating behavior).
- **athlete - partner** (impact on performance).
- **coach - athlete** (this relationship is considered to be particularly crucial).
- **athletes –clubs,** etc.

New tools- new questions

- The increasing availability of massive networked data is revolutionizing the scientific study of a variety of phenomena in all fields of sports.

- A variety of techniques from different disciplines can help to analyze and mine the information in sports networks, study their distribution, their diffusion, their value, and their influence on social and economic outcomes.

- Petabytes of data regarding **human movements**, **transactions**, and **communication patterns** are continuously being generated by everyday technologies (e.g. mobile phones, credit cards, Myspace, LinkedIn, Friendster, Facebook, etc).

- The popularity of these social applications rely on the networking power of communities. Regarding sports, interesting questions can be asked related to the **behavioral dynamics** of sports networks. Answers to these questions can have implications in **marketing, advertising strategies, ticket sales, safety in large sport events**, etc.

Research Questions

- Can we interprete the **small-world phenomenon** by examining the patterns of connections in large-scale sports networks (this linkage pattern can itself explain the small-world results)?

Research Questions

- Using existing patterns of connections in a sport network and a variety of **graph-theoretic** and **statistical techniques**, how can we predict new relationships that will form in the network in the near future?

Research Questions

- What are the **dynamics of information flow** in a sports network? How can we extract **(data mining)** knowledge from this information?

Research Questions

- How can the future of a sports network be predicted from the current state of the network?

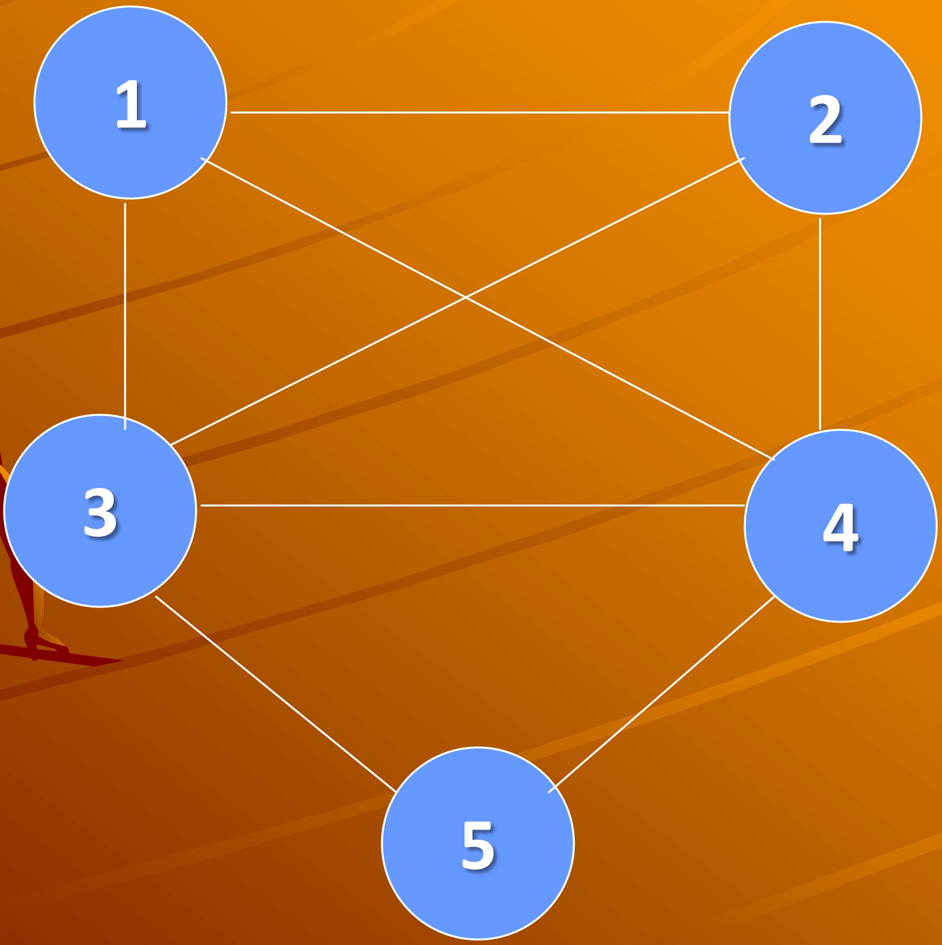
- Prediction is very difficult, especially of the future.

  \- Niels Bohr (1885-1962)

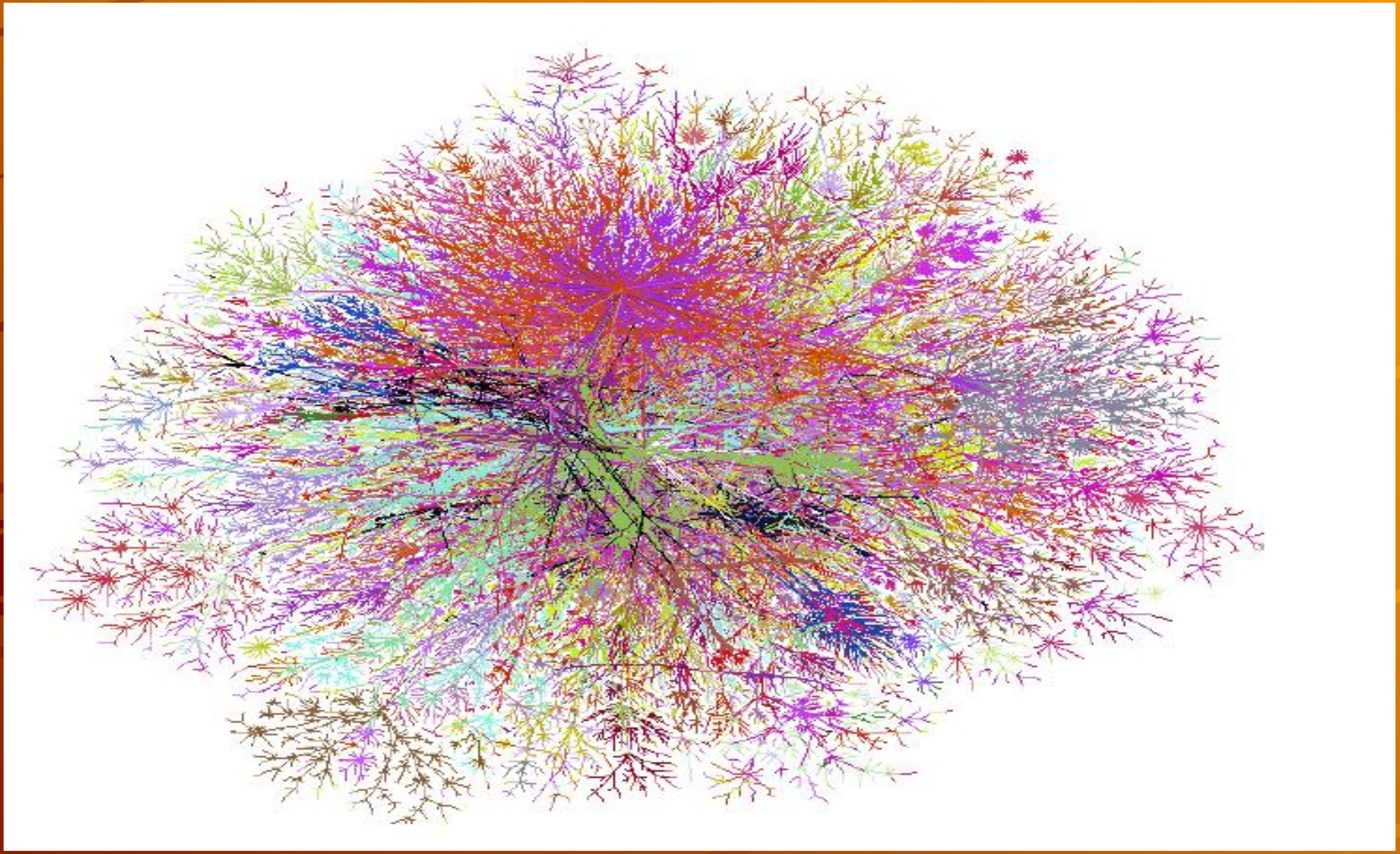
Representing a dataset as a graph

- In many cases, a massive dataset can be represented as a graph
- A graph is a set of dots (vertices) and links (edges)

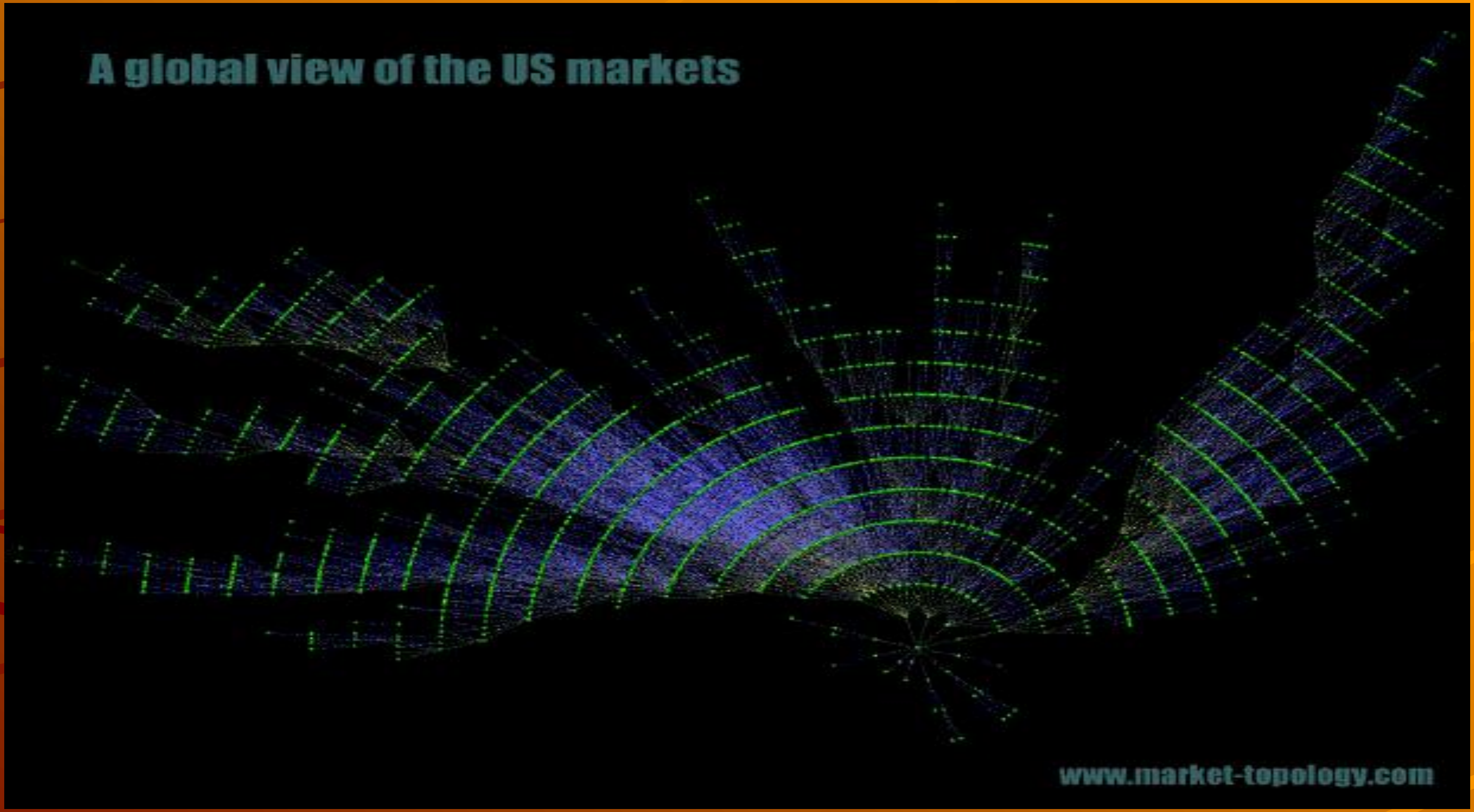
Example of a graph



Examples: Internet Graph



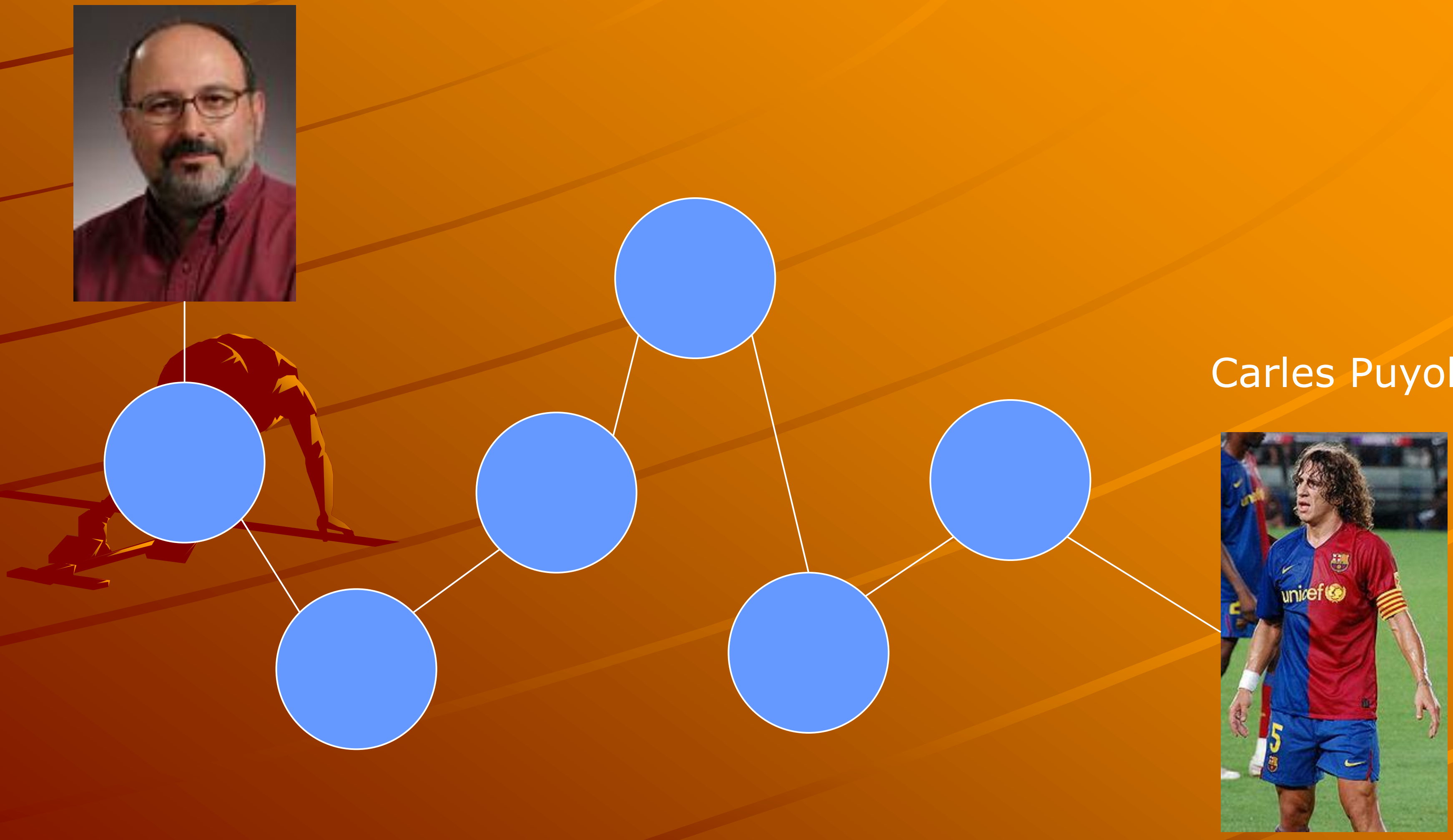
Examples: Financial graphs



Examples: Social networks

- Real people are the vertices
- An edge connects two people if they know each other
- These networks are sparse and clustered, and they have a small diameter
- 6 degrees of separation
- “Small world” hypothesis

Small world hypothesis



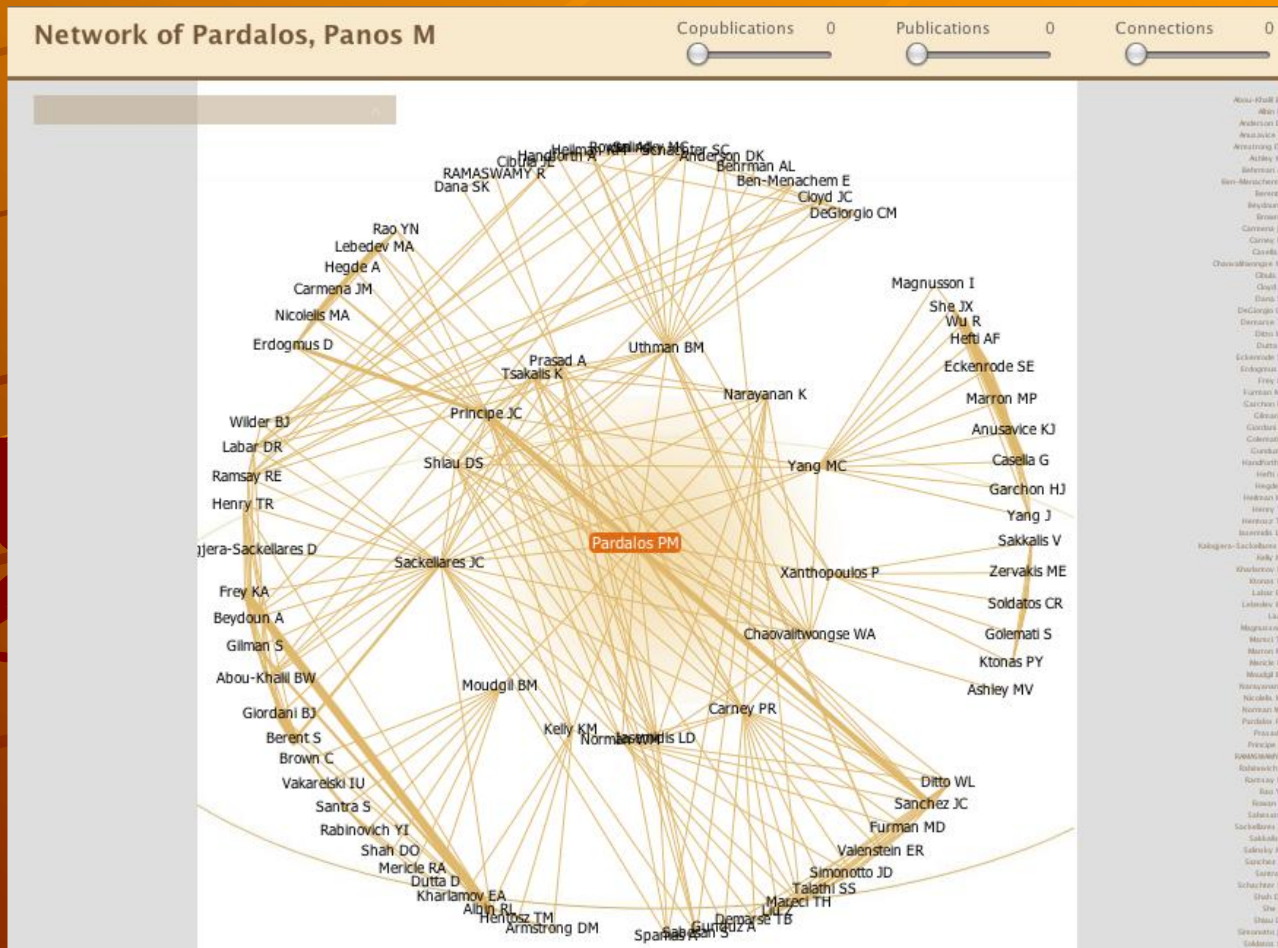
Social networks (cont'd)

- Scientific collaboration graph: *Erdös* number
- e.g. My *Erdös* Number is 2:

Ding-Zhu Du, Xi'an Jiaotong, Ronald L. Graham, Panos M. Pardalos, Peng-Jun Wan, Weili Wu, Wenbo Zhao, "Analysis of greedy approximations with nonsubmodular potential functions", *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms, 167-175, 2008.P.*

Erdös, R. L. Graham, "Packing squares with equal squares", Journal of Combinatorial Theory Series A, 19:1, 119-123, 1975.

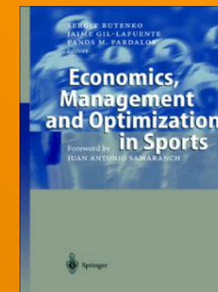
Collaboration Networks



Books

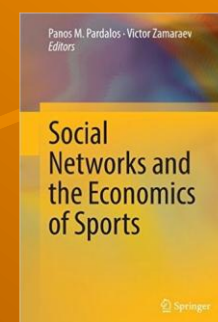
J. Abello, P.M. Pardalos, and M. Resende,

Handbook of Massive Data Sets,Kluwer Academic Publishers, (2002).



S. Butenko, J. Gil-Lafuente, and P.M. Pardalos,
Economics, Management and Optimization in Sports, Springer, (2004).

S. Butenko, J. Gil-Lafuente, and P.M. Pardalos,
Optimal Strategies in Sports Economics and Management, Springer, (2010).

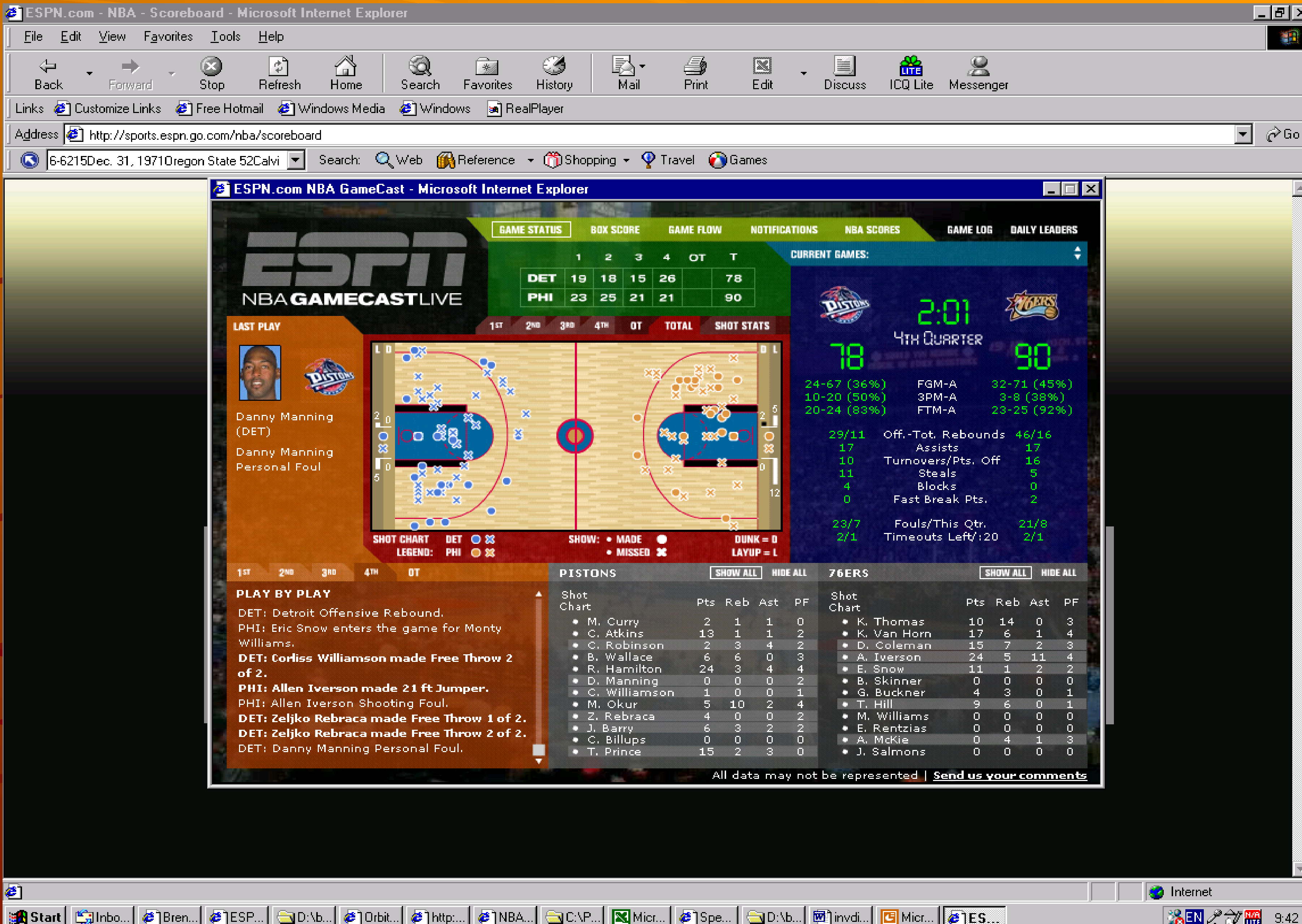


P.M. Pardalos and V. Zamaraev

Social Networks and the Economics of Sports, Springer, (2014).

NBA Graph

- Basketball players are vertices
- An edge connects two vertices if these two players ever played in the same team
- We consider the graph consisting of 404 vertices, i.e. 404 players currently playing in the NBA


ESPN.com - NBA - Scoreboard - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss ICQ Lite Messenger
Links Customize Links Free Hotmail Windows Media Windows RealPlayer
Address http://sports.espn.go.com/nba/scoreboard
Go
Search: Web Reference Shopping Travel Games
ESPN.com NBA GameCast - Microsoft Internet Explorer
GAME STATUS BOX SCORE GAME FLOW NOTIFICATIONS NBA SCORES GAME LOG DAILY LEADERS
NBA GAMECASTLIVE
1 2 3 4 OT T
DET 19 18 15 26 78
PHI 23 25 21 21 90
CURRENT GAMES:
LAST PLAY
Danny Manning (DET)
Danny Manning Personal Foul
1ST 2ND 3RD 4TH OT TOTAL SHOT STATS
2:01
4TH QUARTER
78
90
24-67 (36%) FGM-A 32-71 (45%)
10-20 (50%) 3PM-A 3-8 (38%)
20-24 (83%) FTM-A 23-25 (92%)
29/11 Off.-Tot. Rebounds 46/16
17 Assists 17
10 Turnovers/Pts. Off 16
11 Steals 5
4 Blocks 0
0 Fast Break Pts. 2
23/7 Fouls/This Qtr. 21/8
2/1 Timeouts Left/:20 2/1
SHOT CHART LEGEND: DET PHI
SHOW: MADE MISSED
DUNK = D LAYUP = L
PLAY BY PLAY
DET: Detroit Offensive Rebound.
PHI: Eric Snow enters the game for Monty Williams.
DET: Corliss Williamson made Free Throw 2 of 2.
PHI: Allen Iverson made 21 ft Jumper.
PHI: Allen Iverson Shooting Foul.
DET: Zeljko Rebraca made Free Throw 1 of 2.
DET: Zeljko Rebraca made Free Throw 2 of 2.
DET: Danny Manning Personal Foul.
PISTONS SHOW ALL HIDE ALL
Shot Chart Pts Reb Ast PF
M. Curry 2 1 1 0
C. Atkins 13 1 1 2
C. Robinson 2 3 4 2
B. Wallace 6 6 0 3
R. Hamilton 24 3 4 4
D. Manning 0 0 0 2
C. Williamson 1 0 0 1
M. Okur 5 10 2 4
Z. Rebraca 4 0 0 2
J. Barry 6 3 2 2
C. Billups 0 0 0 0
T. Prince 15 2 3 0
76ERS SHOW ALL HIDE ALL
Shot Chart Pts Reb Ast PF
K. Thomas 10 14 0 3
K. Van Horn 17 6 1 4
D. Coleman 15 7 2 3
A. Iverson 24 5 11 4
E. Snow 11 1 2 2
B. Skinner 0 0 0 0
G. Buckner 4 3 0 1
T. Hill 9 6 0 1
M. Williams 0 0 0 0
E. Rentzias 0 0 0 0
A. McKie 0 4 1 3
J. Salmons 0 0 0 0
All data may not be represented | Send us your comments
Internet
Start 9:42

Structure of the NBA graph

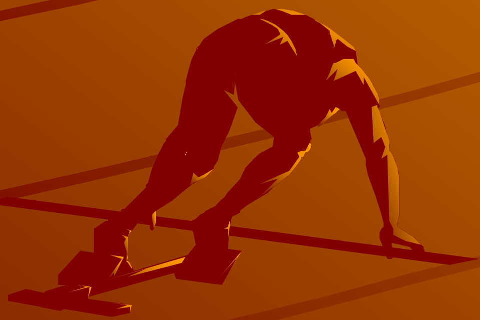
- It is *sparse*: the number of edges is 5492, whereas the maximum possible number of edges is 81406.

So the edge density is 6.75%.

- It is *clustered*: two vertices are more likely to be connected if they have a common neighbor.

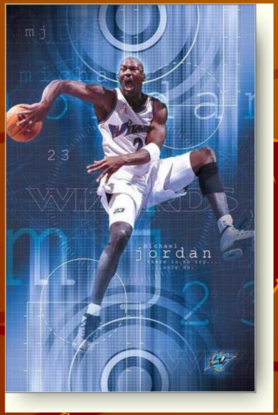
Jordan number

- Jordan number of each vertex (player) is equal to the distance from this vertex to the vertex representing Michael Jordan



Jordan number

0



1

Scottie Pippen

Toni Kukoc

Jerry Stackhouse

2

Shaquille O’Neal

Kobe Bryant

Arvydas Sabonis

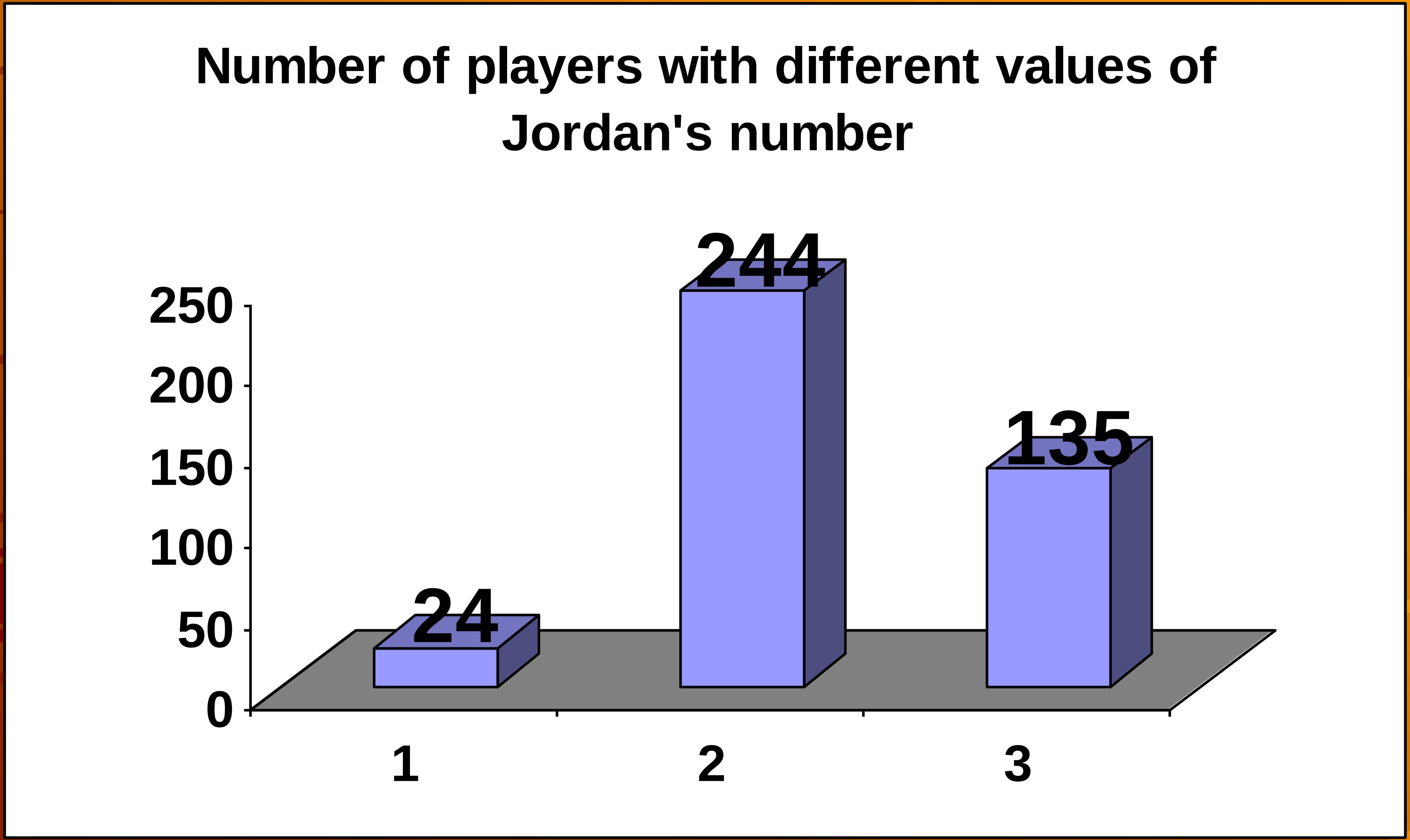
3

Yao Ming

Dirk Nowitzki

Pau Gasol

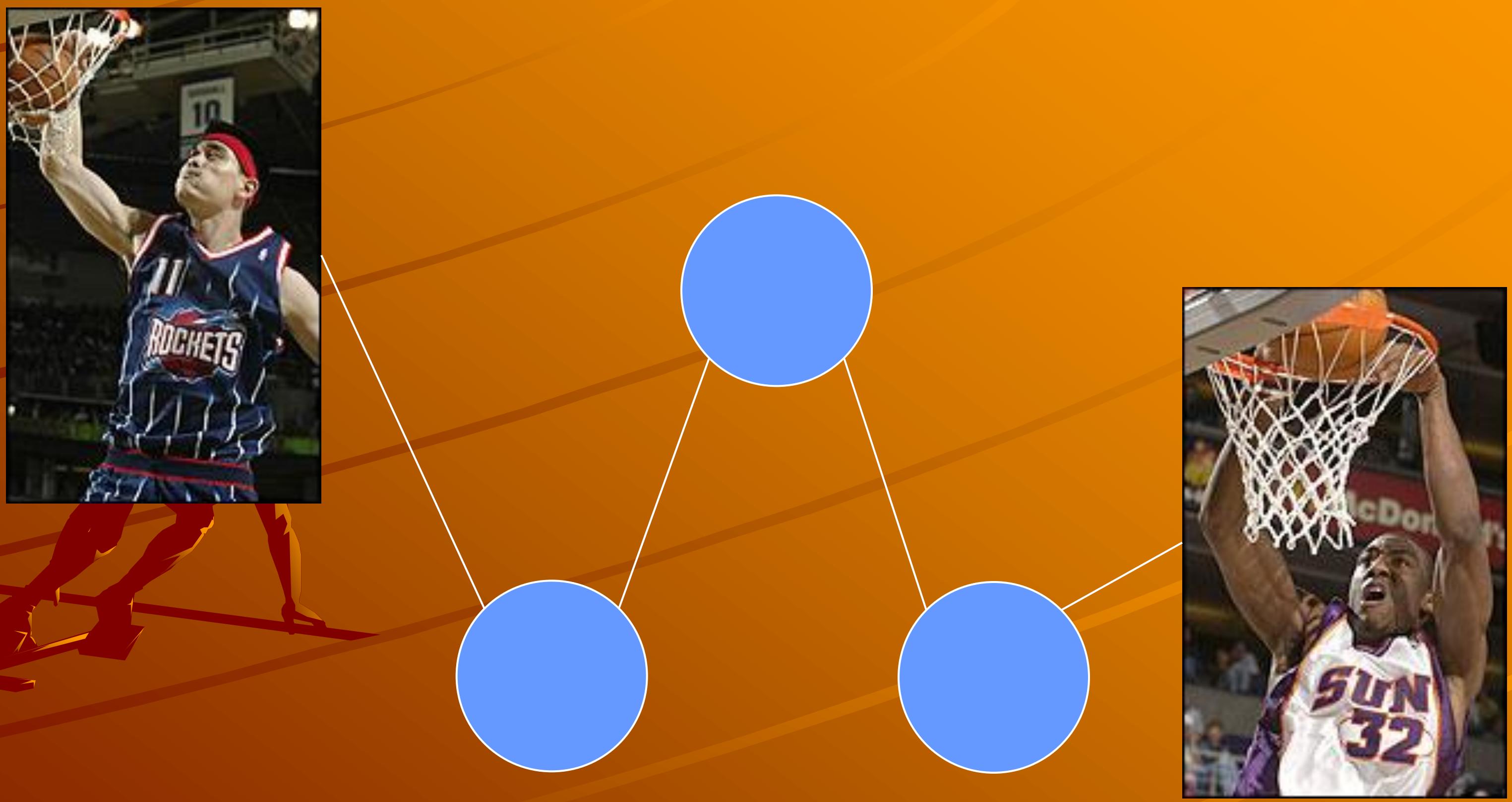
Jordan number



Diameter of the NBA Graph

- The *diameter* of the graph is maximum possible distance (i.e., number of edges) between any pair of vertices.
- The diameter of the NBA graph is equal to 4.
- We can talk about 3 degrees of separation between any two players: NBA is a very small world!

3 degrees of separation in the NBA graph



Degree distribution

- The *degree* of a vertex is the number of edges originating from it

Players with highest degrees in the NBA graph

Jim Jackson, Sacramento Kings
Degree: 68

Corie Blount, Chicago Bulls
Degree: 63

Robert Pack, New Orleans Hornets
Degree: 57

Extensions of this approach

- It is possible to construct a graph representing all European football players
- Since there are a lot of transfers between European football teams each year, this graph is believed to have a small diameter

- V. Boginski, S. Butenko, P.M. Pardalos, O. Prokopyev *Collaboration Networks in Sports*. Economics, Management and Optimization in Sports, (Editors) S. Butenko, J.Gil-Lafuente, P.M. Pardalos, Springer, 265-278 (2003)

The Dutch soccer team as social network (DST)

- In the DST network every node corresponds to a player that has played an official match for the Dutch Soccer Team (data from www.voetbalstats.nl)

Number of nodes 691

Number of links 10450

Density 0.044

The DST network is connected, i.e. between any two players a path exists.

The Dutch soccer team as social network

- The DST network is a small world network, because the average distance between players is small (4.49), while the clustering coefficient is high (0.75).
- The player with the most coplayers is Harry Dénis, who played together with 117 others.

The Dutch soccer team as social network

- Of all players Harry Dénis has the lowest clustering coefficient, i.e. he is the player whose co-players are the least mutually connected.
- The diameter of the DST network is 11, i.e. the longest shortest path has length 11.
- The most central player in the DST network is Roel Wiersma.

The Dutch Soccer Team as a Social Network (by Robert Kooij, Almerima Jamakovic, Frank van Kesteren, Tim de Koning,Ildiko Theisler, and Pim Veldhoven)

Brazilian Soccer Players network (BSP)

Number of nodes 13,411
Number of links 315,566

This a bipartite network that contains two types of vertices: one set is created from the 127 different Brazilian soccer clubs and the other set is created from the 13,411 soccer players who have played for those clubs between the years 1971 and 2002.

Brazilian Soccer Players network (BSP)

- In this network, whenever a soccer player has been employed by a certain club they are connected by an edge. It has been shown that the probability that a Brazilian soccer player has worked at *N* clubs or played *M* games shows an exponential decay while the probability that he has scored *G* goals is **power law**.

- In addition, they consider a network composed only by the soccer players. If two players were at the same team at the same time, then they will be connected by an edge.

Brazilian Soccer Players network (BSP)

- The BSP network is a **small-world network** with 3.29 degrees of separation between the Brazilian soccer players.

- What about the time evolution of the BSP network?

1. the player's professional life is turning longer and/or the players transfer rate between teams is growing up.
2. The clustering coefficient is a time decreasing function. This may reflect the players transfer rate between national teams and the exodus of the best Brazilian players to foreigner teams (which has increased, particularly, in the last decades).
3. The BSP network is becoming more assortative with time. This seems to indicate the existence of a growing segregationist pattern, where the players transfer occurs, preferentially, between teams of the same size.

Details of this study can be found in: Onody, R.N. and de Castro, P.A. (2004).Complex network study of Brazilian soccer players. Physical Review E, 70(3): 037103.

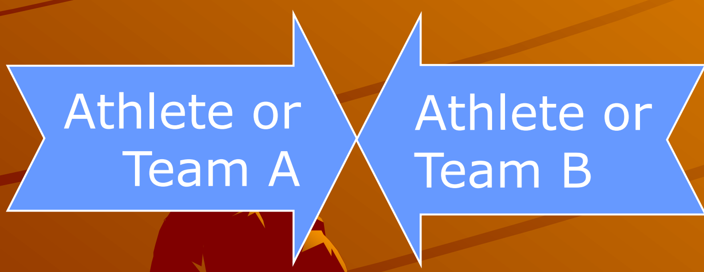
MULTI-CONTESTANT RACES

Introduction of our problem

- **Predictions of sports games** have been recognized as an important area of study for its economic significance.
- The majority of models for such games **cover two-player games** and the resulting championships or study individual players or teams and their resulting comparative position.
- However, many sports involve ***race-type multi-contestant games***, which are more complex in modeling.
- Our work presents a model based exponentially distributed racing times as a first approximation of the problem.

Common problems

The most studied problem:
two-player game



The most studied outcomes:
one wins / other loses
Sometimes **a tie**

Predictions about:

- Individual game outcomes
- Tournaments
- Championships

- Athletes scoring
- Athlete ranking

Multi contestant races

Many athletes participating in a race simultaneously

An individual performance (usually time) determines the winner

The individual scores are ranked and the best is the 1st winner, other positions may also count

The problem becomes a problem of **ranking** random variables and **order statistics**

Challenges

Study the problem as a problem of ranking (order statistics)

Easy IF:

- The **times distribution** is known and the individual performances are **independent**

or

- the **joint distribution** is known

We need consistent historical data about the individual performances (e.g the times) in the games

BUT...

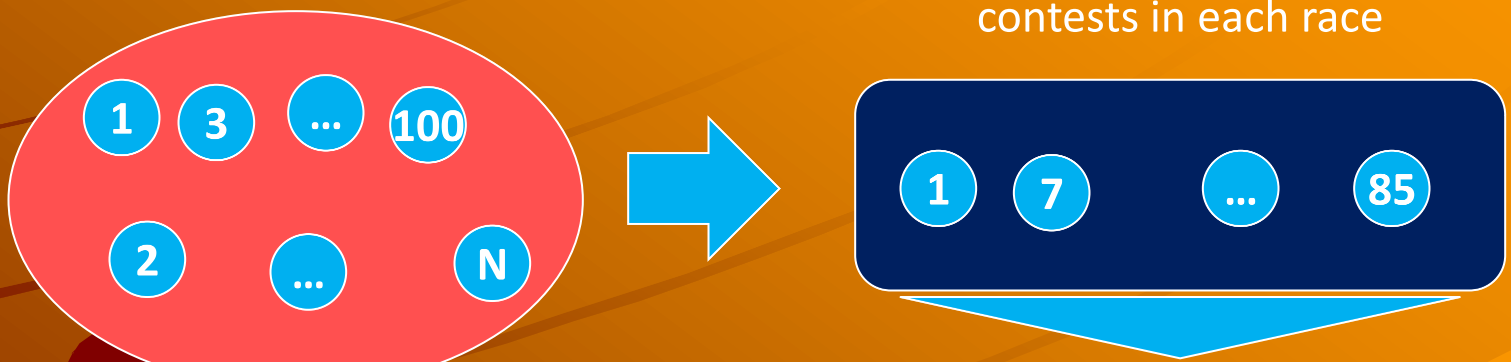
… but

- Performances cannot always be considered independent. The make-up of the contestants in the race influences individual performance
- We do not always know the individual performance measures' size, just the ranks
- Not all athletes have contested against each other in races
- The number of race data are uneven in the sense that we have different number of data for races among different athletes
- Even when we know them they are not immediately comparable because they have been attained under different conditions (weather, altitude etc.)

Definition of our problem

Large pool of *N* athletes

Fixed number of athletes contests in each race



Results come as ranks of the athletes in each race – their performance measures are not available

We want to predict the likelihood of an outcome for a race of athletes from the pool

Our approaches (1)

1. Assume independent exponentially distributed times:
   - then the problem is one of studying the order statistics of the exponential distribution
   - we don’t really need to know the times as we can estimate the parameters relative to each other
   - we can find Maximum Likelihood or Least Squares estimates => we need to study carefully the effect of uneven data on the confidence level

Our approaches (2)

2. Drop the assumption of independence and use a joint distribution function:
   – we have identified as appropriate an extension of a bivariate presented by Gumbel in 1960

$$f(x_1, x_2, \ldots, x_m) = \lambda^* e^{-\sum_{i=1}^{m} \lambda_i x_i} \left[1 + \sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} \frac{\delta_{ij}}{\delta^*} (2e^{-\lambda_i x_i} - 1)(2e^{-\lambda_j x_j} - 1) \right]$$

   – we can use data for estimation but there is a difficulty for athletes that have not met in a race before. There we can assume independence

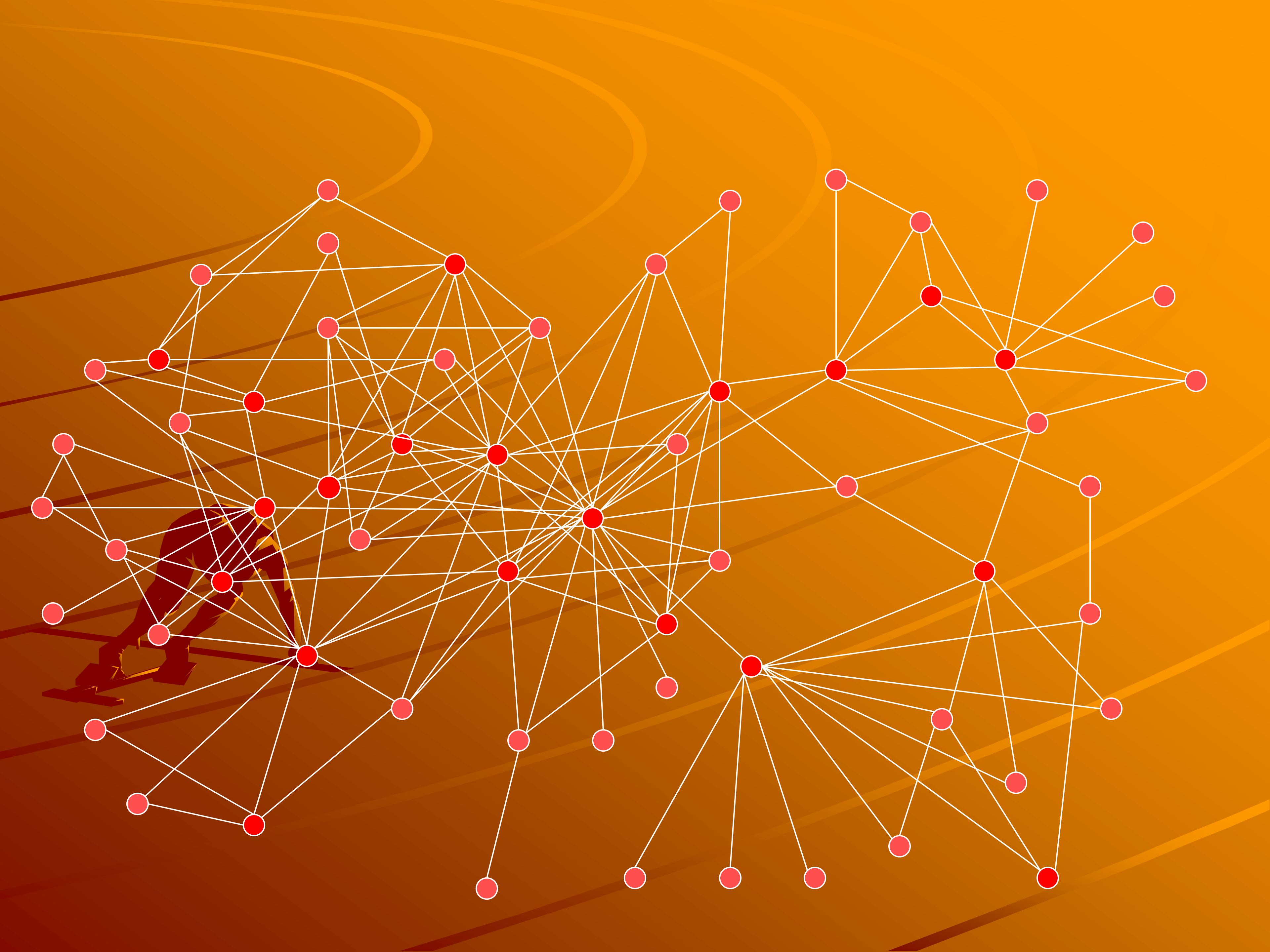
Our approaches (3)

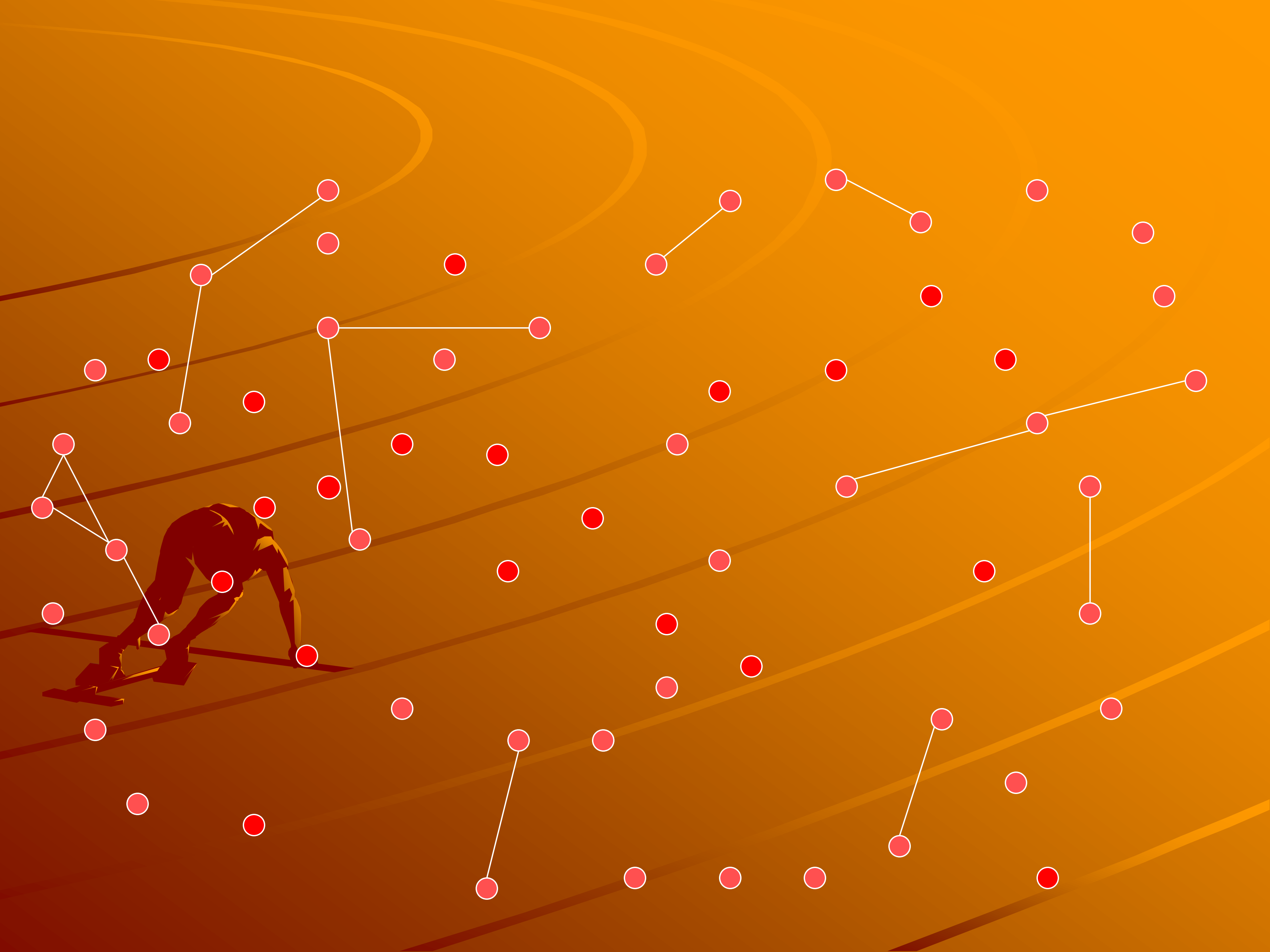
1. When we have athletes that have not met in any race, we can establish their relative dominance using a network approach based on relative results with other athletes that have contested against the two athletes compared.

Next steps

- Apply known estimation techniques for the parameters of the joint distribution
- Try to understand the level of significance of the estimation and subsequently the prediction
- Test on past to see fitness of the approach

CRITICAL NODE (AND EDGE) DETECTION (CND) PROBLEMS





Problem Definition

- Given a graph G = (V,E) and an integer k
- Goal is to detect (delete) a set $|A| \leq k$ of critical nodes, or nodes whose deletion results in maximum pairwise disconnectivity
- Disconnectivity → MAX components subject to MIN difference in cardinality

Problem Definition

- Decision Version: K-CNP
- Input: Undirected graph *G = (V,E)* and integer k
- Question: Is there a set *M*, where *M* is the set of all maximal connected components of *G* obtained by deleting *k* nodes or less, such that

$$\sum_{\forall i \in M} \frac{\sigma_i(\sigma_i - 1)}{2} \leq K$$

where σ_i is the cardinality of component *i*, for all *i* in *M*?

- The summation term expresses the pairwise connectivity in the graph

Formulation

- Let $u_{i,j} = 1$, if i and j are in the same component of G(V \ A), and 0 otherwise.
- Let $v_i = 1$, if node i is deleted in the optimal solution, and 0 otherwise.
- We can formulate the CNP as the following integer linear program.

Formulation

$$
\begin{aligned}
\text{(CNP-1)} \quad \text{Minimize} \quad & \sum_{i,j \in V} u_{ij} \\
\text{s.t.} \quad & \\
& u_{ij} + v_i + v_j \geq 1, \ \forall \ (i,j) \in E, \\
& u_{ij} + u_{jk} - u_{ki} \leq 1, \ \forall \ (i,j,k) \in V, \\
& u_{ij} - u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V, \\
& -u_{ij} + u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V, \\
& \sum_{i \in V} v_i \leq k, \\
& u_{ij} \in \{0,1\}, \ \forall \ i,j \in V, \\
& v_i \in \{0,1\}, \ \forall \ i \in V.
\end{aligned}
$$

Formulation

(CNP-1) Minimize $\sum_{i,j \in V} u_{ij}$

s.t.

$$u_{ij} + v_i + v_j \geq 1, \ \forall \ (i, j) \in E,$$

$$u_{ij} + u_{jk} - u_{ki} \leq 1, \ \forall \ (i, j, k) \in V,$$

$$u_{ij} - u_{jk} + u_{ki} \leq 1, \ \forall \ (i, j, k) \in V,$$

$$-u_{ij} + u_{jk} + u_{ki} \leq 1, \ \forall \ (i, j, k) \in V,$$

$$\sum_{i \in V} v_i \leq k,$$

$$u_{ij} \in \{0, 1\}, \ \forall \ i, j \in V,$$

$$v_i \in \{0, 1\}, \ \forall \ i \in V.$$

If i and j in different components and there is an edge between them, at least one must be deleted

Formulation

(CNP-1) Minimize $\sum_{i,j \in V} u_{ij}$

s.t.

$$u_{ij} + v_i + v_j \geq 1, \ \forall \ (i,j) \in E,$$

$$u_{ij} + u_{jk} - u_{ki} \leq 1, \ \forall \ (i,j,k) \in V,$$

$$u_{ij} - u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V,$$

$$-u_{ij} + u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V,$$

$$\sum_{i \in V} v_i \leq k,$$

$$u_{ij} \in \{0,1\}, \ \forall \ i,j \in V,$$

$$v_i \in \{0,1\}, \ \forall \ i \in V.$$

Number of nodes deleted is at most k.

Formulation

$$
\begin{aligned}
\text{(CNP-1)} \quad \text{Minimize} \quad & \sum_{i,j \in V} u_{ij} \\
\text{s.t.} \quad & \\
& u_{ij} + v_i + v_j \geq 1, \ \forall \ (i,j) \in E, \\
& u_{ij} + u_{jk} - u_{ki} \leq 1, \ \forall \ (i,j,k) \in V, \\
& u_{ij} - u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V, \\
& -u_{ij} + u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V, \\
& \sum_{i \in V} v_i \leq k, \\
& u_{ij} \in \{0,1\}, \ \forall \ i,j \in V, \\
& v_i \in \{0,1\}, \ \forall \ i \in V.
\end{aligned}
$$

For all triplets (i,j,k), if (i,j) in same comp and (j,k) in same comp, then (i,k) in same comp.

Formulation

(CNP-1) Minimize $\sum_{i,j \in V} u_{ij}$

s.t.

$$u_{ij} + v_i + v_j \geq 1, \ \forall \ (i,j) \in E,$$

$$u_{ij} + u_{jk} - u_{ki} \leq 1, \ \forall \ (i,j,k) \in V,$$

$$u_{ij} - u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V,$$

$$-u_{ij} + u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V,$$

$$\sum_{i \in V} v_i \leq k,$$

$$u_{ij} \in \{0,1\}, \ \forall \ i,j \in V,$$

$$v_i \in \{0,1\}, \ \forall \ i \in V.$$

Can combine into one constraint for a simpler model

Formulation

$$\textbf{(CNP-2)} \quad \text{Minimize} \quad \sum_{i,j \in V} u_{ij}$$

s.t.

$$u_{ij} + v_i + v_j \geq 1, \ \forall \ (i,j) \in E,$$

$$u_{ij} + u_{jk} + u_{ki} \neq 2, \ \forall \ (i,j,k) \in V,$$

$$\sum_{i \in V} v_i \leq k,$$

$$u_{ij} \in \{0,1\}, \ \forall \ i,j \in V,$$

$$v_i \in \{0,1\}, \ \forall \ i \in V,$$

$$u_{ij} + u_{jk} - u_{ki} \leq 1, \ \forall \ (i,j,k) \in V,$$

$$u_{ij} - u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V,$$

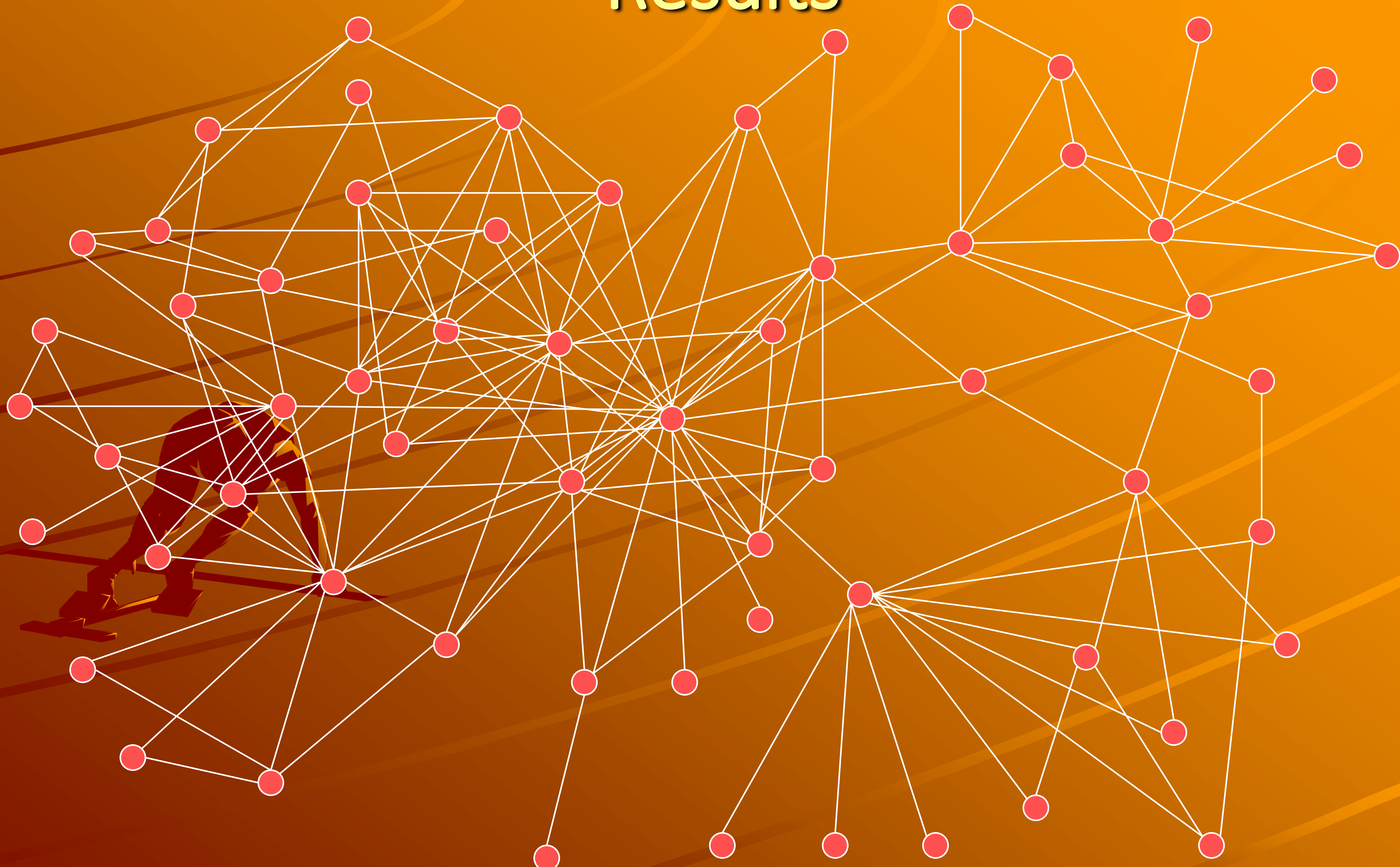
$$-u_{ij} + u_{jk} + u_{ki} \leq 1, \ \forall \ (i,j,k) \in V,$$

Exact algorithms and Heuristics

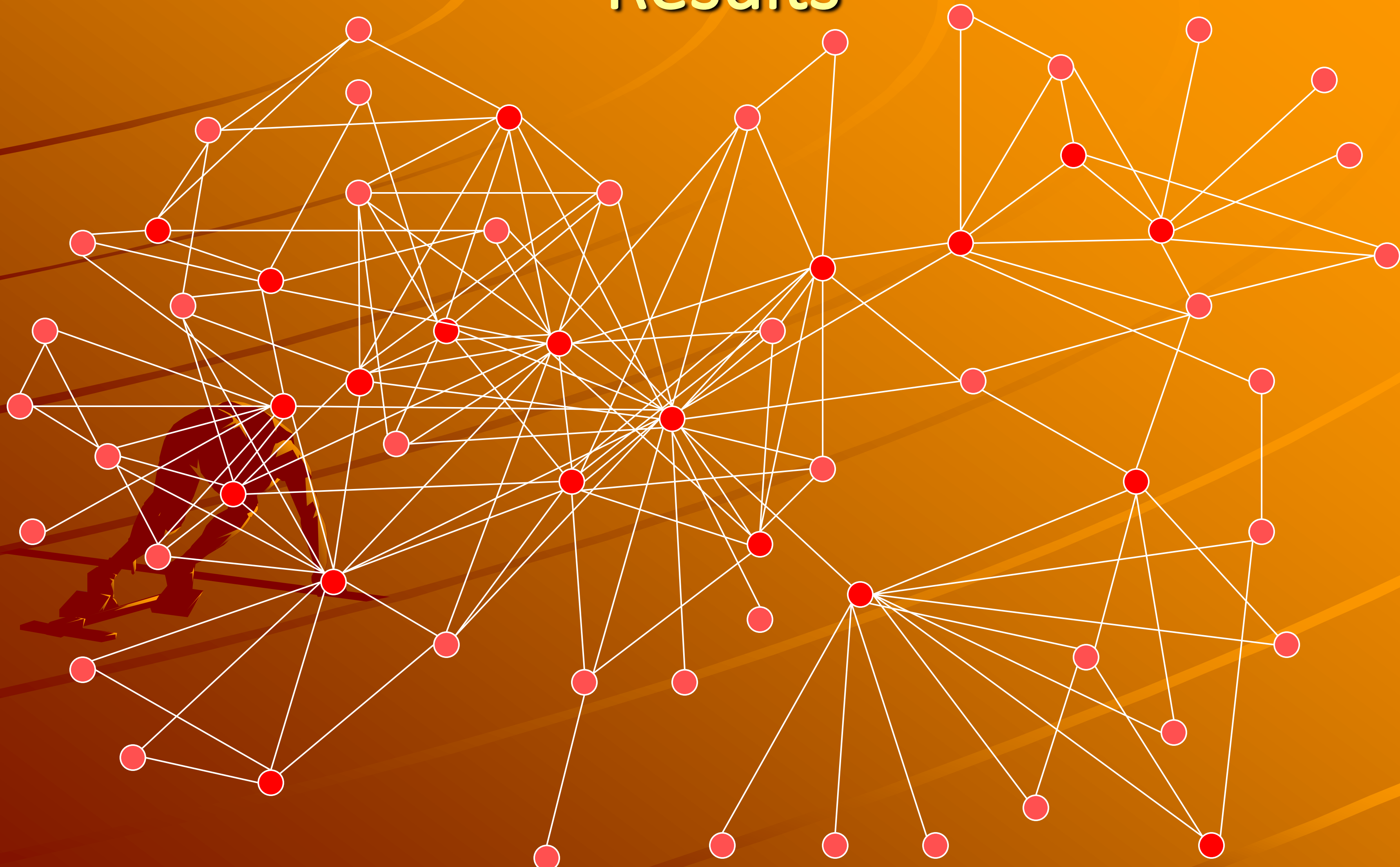
- For large scale problems we developed heuristics based on max independent set problem, GRASP, and genetic algorithms.

- A. Arulselvan, C. Commander, L. Elefteriadou, P.M. Pardalos. *Detecting Critical Nodes in Sparse Graphs*, *Computers and Operations Research*, 36(7), 2193-2200, 2009.

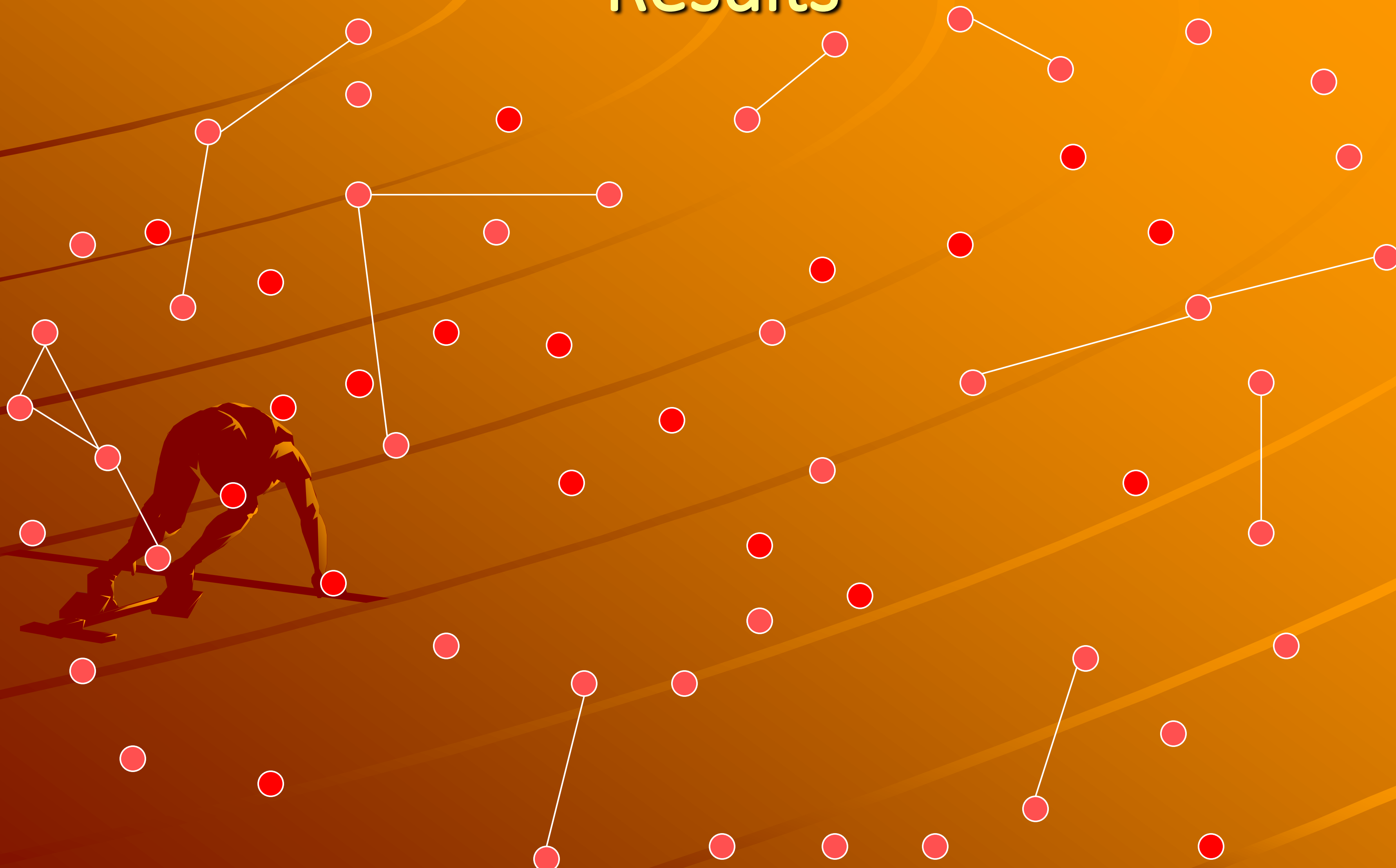
Results



Results



Results



Concluding remarks and questions

- In this talk, we have surveyed several social collaboration networks in sports and discussed their structure. All such networks, which consist of players only proved to be small worlds. One might assume that networks which can be constructed in the same way for other team games or leagues have the similar structure.
- There are many interesting questions related to the behavioral dynamics of sports networks.

Concluding remarks and questions

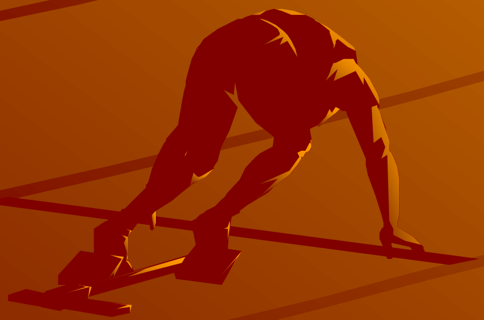
- For example:
  - Using the existing patterns of connections in a sports network and a variety of graph-theoretic and statistical techniques, how can one predict a new relationships that will form in the network in the near future?
  - What are the dynamics of the information flow in a sports network? How can one extract knowledge from this information?
  - How can the future of a sports network be predicted from the current state of the network?

Concluding remarks and questions

- For example:
  - Using the existing patterns of connections in a sports network and a variety of graph-theoretic and statistical techniques, how can one predict a new relationships that will form in the network in the near future?
  - What are the dynamics of the information flow in a sports network? How can one extract knowledge from this information?
  - How can the future of a sports network be predicted from the current state of the network?

Concluding remarks and questions

- Answers to these questions may have implications in management, marketing, advertising strategies, ticket sales, safety in large sports events, etc



«νοῦς ὑγιής ἐν σώματι ὑγιεῖ»

(“A healthy mind in a healthy body”)

- based on Hippocrates (460 BC-337 BC)

You can't build a reputation on what you are going to do.

- Henry Ford (1863-1947)

Fantasy League?

Did you Analyze Your Team's Network First?

Panos M. Pardalos
Distinguished Professor of Industrial and Systems Engineering

Sports have been an integral part of human culture since ancient times and play a key role in the economy, politics, and lifestyle of any country.

Many sports, including football, baseball, basketball, track and field, hockey, tennis, and golf, and their associated industries, rely on decision-making tools from a wide spectrum of techniques in economics, networks, and optimization.

The ancient Greeks spoke vividly about the human condition, with phrases such as a "healthy mind in a healthy body." The thought of exercising the human spirit along with the human body was established very early in the Greek world, thus instilling the concept of physical and mental harmony of the human being.

In the Greek states, social processes aimed towards intellectual development and physical perfection eventually culminated in the institutionalization of the Olympic Games, which graced the ancient world with the ideal of fair play for over a thousand years. This tradition inspired people to revive the Modern Olympics, thus promoting the ideal of fair play in the modern world. This resulted in the amazing proliferation of sports in our time, while also contributing to a giant economic boom spurred by the extended activities around sports and the Olympics.

Today's sports industry is complex and impacts several economic markets, such as television, advertising, clothing, and manufacturing. In addition, sports are characterized by a unique need for competitive balance. As early as 1964, the economist Walter Neale stated the "Louis-Schmeling paradox" in that better profits could be made from a better product, which in boxing, meant two strong fighters. Louis could not have made it without a strong Schmeling.

It is clear that in most businesses the ideal market position of a company is a monopoly. But in sports, it is much different. Given the paradox, a pure monopoly would be a disaster. Fans want to see a competitive balance among teams in order to keep their interest. (Neale referred to this as "inverted joint products.")

Systems engineering tools can be used to study many issues in sports. For example, social network analysis deals with the interactions between individuals by considering them as nodes of a network whereas their relations are mapped as network edges. The study of such structures lies at the intersection of different disciplines of research, including economics, sociology and graph theory. In practice, many kinds of networks have been studied, including friendship networks, scientific co-authorship networks, film collaboration networks, disease spreading networks, and urban growth networks.

We have also studied social networks arising in sports. The structure of sports networks, as well as properties of their dynamics, can be useful in the management, economics and marketing of sports.

Social networks arise naturally in sports, especially team sports. For example, such networks may be based on the various relationships: athlete-athlete, coach-athlete, athlete-clubs, etc. In addition to a single player analysis, the analysis of sports networks may be used to investigate patterns of social relations between team members, as well as to explore the behavioral dynamics of groups of teammates. This information may help coaches, players, managers and other club members in making decisions.

As an example, we (Boginski et. al, 2004) considered the National Basketball Association (NBA) from the perspective of social networks. We constructed and studied a social network graph of players in the NBA. It turned out that the properties of this graph are similar to the properties of other social networks.

The vertices of the NBA graph represent all the basketball players who played during the 2002-03 season. An edge joins two given players if they ever played in the same team. The constructed graph has 404 nodes and 5492 edges between them. Thus, the edge density of this graph is rather small: 5492/81406 = 6.75%.

Based on the definition, one can state some obvious structural properties of the NBA graph. For example, players from the same team form a clique in the graph. Also, since many players change teams, there exist links between players (vertices) from different teams (cliques). Clearly, the same structure is inherent for all collaboration networks.

The considered NBA graph is connected, due to the fact that the number of players in a basketball team is relatively small and players transfer between different teams frequently. Furthermore, we observed that this graph has a 'small-world' topology. That is, the maximum distance between all pairs of vertices (graph diameter) in the NBA graph is equal to 4, implying that the NBA is a very small world.

Analogous to the Erdos number for the collaboration graph of mathematicians and the Kevin Bacon number for a graph of Hollywood actors, the "Jordan number" was introduced for the NBA graph. A player's Jordan number is defined as the distance in the NBA graph between the vertex corresponding to that player and the vertex corresponding to the former Chicago Bulls player, Michael Jordan, arguably the greatest basketball player ever.

Despite the fact that Michael Jordan only played for two teams throughout his career, and thus had relatively few "collaborators," most players (268) in the studied NBA graph have a Jordan number of either one or two; the maximum Jordan number is only three. This means that all players are connected to Jordan through at most two vertices, which again confirms that NBA is a 'small world'.

Similar collaboration networks have been considered for soccer and other types of sports. All such networks consisting of players have proved to be 'small worlds'. One might suppose that networks which can be constructed in the same way for other team games or leagues have similar structure.

There are many interesting questions related to the behavioral dynamics of sports networks. For example:

- Using the existing patterns of connections in a sports network and a variety of graph-theoretic and statistical techniques, how can one predict new relationships that will form in the network in the near future?
- What are the dynamics of the information flow in a sports network? How can one extract knowledge from this information?
- How can the future of a sports network be predicted from the current state of the network?

Answers to these questions may have implications in management, marketing, advertising strategies, ticket sales, and safety in large sporting events.

For further reading

V. Boginski, S. Butenko, P. M. Pardalos, and O. Prokopyev. Collaboration networks in sports. In S. Butenko, J. Gil-Lafuente, and P. M. Pardalos, editors, *Economics, Management and Optimization in Sports*, pages 265-277. Springer, 2004.

S. Butenko, J. Gil-Lafuente, P.M. Pardalos (eds.), *Economics, Management and Optimization in Sports*, Springer, 2004.

S. Butenko, J. Gil-Lafuente, P.M. Pardalos (eds.), *Optimal Strategies in Sports Economics and Management*, Springer, 2010.

Thai, M.T. and P.M. Pardalos (eds.) *Handbook of Optimization in Complex Networks: Communication and Social Networks*, Springer, 2011.