Social Networks in Sports

V. Boginski[1], S. Butenko[2], P. M. Pardalos[1], and O. Prokopyev[1]

[1] Industrial and Systems Engineering Department, University of Florida, Gainesville, FL
{vb,pardalos,oap4ripe}@ufl.edu
[2] Industrial Engineering Department, Texas A&M University, College Station, TX
butenko@tamu.edu

1 Introduction

We live in the world of information, where huge amounts of data of diverse nature and origin arise in various spheres of life, including sports. To get useful information from this data, one should apply special techniques of summarizing and visualizing the information contained in a certain dataset. In many practical situations, a real-life dataset can be represented as a large *graph* (*network*) - a structure that can be easily understood and visualized [3]. A graph is a set of vertices (dots) and edges (links) connecting them. When a dataset is represented as a graph, certain attributes are associated with the vertices and edges of the corresponding graph. These attributes may contain specific information characterizing the given application, which often provides a new insight into the internal structure and patterns of the data.

The examples of representing real-life datasets as graphs come from diverse areas. Among these examples, one can mention the Web graph that naturally represents the World-Wide Web [5], the Call graph arising in the telecommunications traffic data [1, 2, 6]; the Market graph reflecting the structure of the stock market [4], and metabolic networks arising in biology [7].

In this chapter, we will discuss one of the most interesting real-life graph applications – so-called “social networks” where the vertices are real people [6, 10]. The main idea of this approach is to consider the “acquaintanceship graph” connecting the entire human population. In this graph, an edge connects two given vertices if the corresponding two persons know each other.

Social networks are associated with a famous “small-world” hypothesis, which claims that despite the large number of vertices, the distance between any two vertices (or, the *diameter* of the graph) is small. More specifically, the idea of “six degrees of separation” has been introduced. It states that any two persons in the world are linked with each other through a sequence of at most six people [6, 10, 11].

Clearly, one cannot verify this hypothesis for the graph incorporating more than 6 billion people living on the Earth, however, smaller subgraphs of the acquaintanceship graph connecting certain groups of people can be investigated in detail. One of the most well-known graphs of this type is the *scientific collaboration graph* reflecting the information about the joint works between all scientists. Two vertices are connected by an edge if the corresponding two scientists have a joint research paper. Another graph of this type is known as the "*Hollywood graph*": it links all the movie actors, and an edge connects two actors if they ever appeared in the same movie [9]. Well-known concepts associated with these graphs are so-called "Erdös number" (in the scientific collaboration graph) and "Bacon number" (in the Hollywood graph), which are assigned to every vertex and characterize the distance from this vertex to the vertex denoting the "center" of the graph. In the collaboration graph, the central vertex corresponds to the famous graph theoretician Paul Erdös, whereas in the Hollywood graph the same position is assigned to Kevin Bacon.

In this chapter, we discuss graphs of a similar type arising in sports, that represent the players' "collaboration". In these graphs, the players are the vertices, and an edge is added to the graph if the corresponding two players ever played together in the same team. One of the examples of this type of graphs is the graph representing baseball players [8]. For any two baseball players who ever played in the Major League Baseball(MLB), a path connecting them can be found in this graph.

As another instance of social networks in sports, we study the "NBA graph" where the vertices represent all the basketball players who are currently playing in the NBA. We apply standard graph-theoretical algorithms for investigating the properties of this graph, such as its connectivity and diameter (*i.e.*, the maximum distance between all pairs of vertices in the graph). As we will see later in the chapter, this study also confirms the "small-world hypothesis". Moreover, we introduce a distance measure in the NBA graph similar to the Erdös number and the Bacon number. The central role in this graph is given to Michael Jordan, the greatest basketball player of all times, and we refer to this measure as the *Jordan number*.

The rest of the chapter is organized as follows. In Section 2, we provide several basic definitions from graph theory, in order to familiarize the reader with standard properties of graphs. Section 3 describes the examples of social networks (scientific collaboration graph, Hollywood graph and baseball graph) in more detail. Section 4 presents various properties of the NBA graph. Finally, Section 5 concludes the discussion.

2 Graph Theory Basics

To give a brief introduction to graph theory, we introduce several basic definitions and notations. Let $G = (V, E)$ be an undirected graph with the set of n vertices V and the set of edges E.

The graph $G = (V, E)$ is *connected* if there is a path from any vertex to any vertex in the set V. If the graph is disconnected, it can be decomposed into several connected subgraphs, which are referred to as the *connected components* of G.

The *distance* between two vertices is the number of edges in the shortest path between them (it is equal to infinity for vertices representing different connected components). The *diameter* of a graph G is usually defined as the maximal distance between pairs of vertices of G. Note, that in the case of a disconnected graph the usual definition of the diameter would result in the infinite diameter, therefore the following definition is in order. By the diameter of a disconnected graph we mean the maximum finite shortest path length in the graph (which is the same as the largest of diameters of the graph's connected components).

The *degree* of a vertex is the number of edges emanating from it.

The *edge density* the graph is defined as a ratio of its number of edges to the maximum possible number of edges. One can easily check that the maximum possible number of edges can be calculated as $n(n-1)/2$ (where n is the number of vertices).

It is also useful to define certain special formations in a graph, which are referred to as *cliques* and *independent sets*. A *clique* in a graph is a set of completely interconnected vertices (*a complete subgraph*), and an *independent set* is a set of vertices without connections.

3 Examples of Social Networks

In this section, we give a more detailed description of the examples of social networks mentioned in the introduction – the scientific collaboration graph, the Hollywood graph, and the baseball graph.

3.1 Scientific Collaboration Graph and Erdös Number

As it was mentioned above, the vertices of the scientific collaboration graph are scientists, and the edges in this graph connect the scientists who have ever collaborated with each other (*i.e.*, had a joint paper). In order to measure the distances in this graph, the "central vertex" is introduced. This vertex corresponds to Paul Erdös, the father of the theory of random graphs. This vertex is assigned *Erdös number* equal to 0. For all other vertices in the graph, the Erdös number is defined as the distance (i.e., the shortest path length) from the central vertex. For example, those scientists who had a joint paper with Erdös have Erdös number 1, those who did not collaborate with Erdös, but collaborated with Erdös' collaborators have Erdös number 2, *etc.*

Following this logic, one can construct the *connected component* of the collaboration graph with "concentric circles", which would incorporate almost

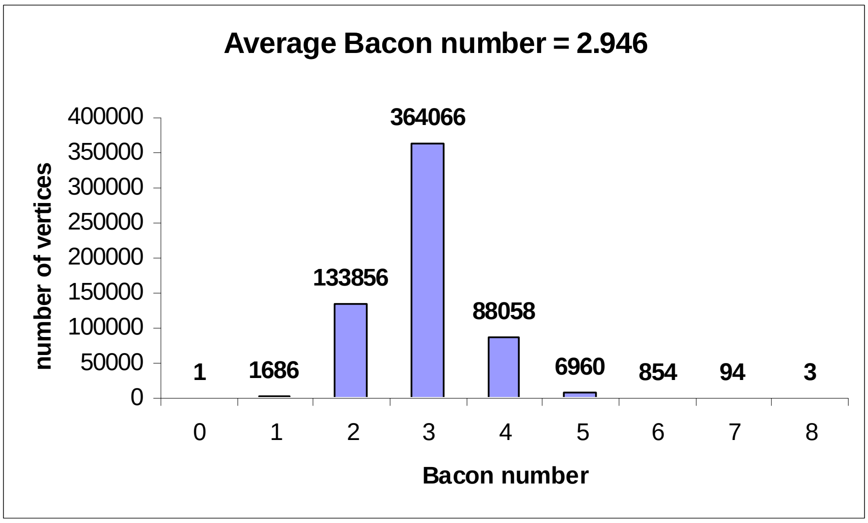

Fig. 1. Number of vertices in the Hollywood graph with different values of Bacon number. Average Bacon number = 2.946.

all scientists in the world, except those who never collaborate with anybody. This connected component is expected to have a relatively small diameter.

The idea of constructing collaboration graphs encompassing people in different areas gave a rise to several other applications. Next, we discuss the Hollywood graph and the baseball graph, where the number of vertices is significantly smaller than in the scientific collaboration graph, which allows one to study their structure in more detail.

3.2 Hollywood Graph and Bacon Number

The Hollywood graph is constructed using the same principles as the scientific collaboration graph, however, the number of Hollywood actors is much smaller than the number of scientists, therefore, one can investigate the characteristics of every vertex in this graph. This information is maintained at the "Oracle of Bacon" website [9]. The most recent Hollywood graph contains 595,578 vertices (actors). The central vertex in this graph represents the famous actor Kevin Bacon, and this vertex obviously has Bacon number 0. Since the number of vertices in this graph is small enough, one can explicitly calculate the Bacon number for every actor. It turns out that most of the actors have Bacon numbers equal to 2 or 3, and the maximum possible Bacon number is equal to 8, which is the case only for 3 vertices.

The distribution of Bacon numbers in the Hollywood graph is shown in Figure 1. The average Bacon number (i.e., the average path length from a given actor to Bacon) is equal to 2.946. As one can see, both the average and

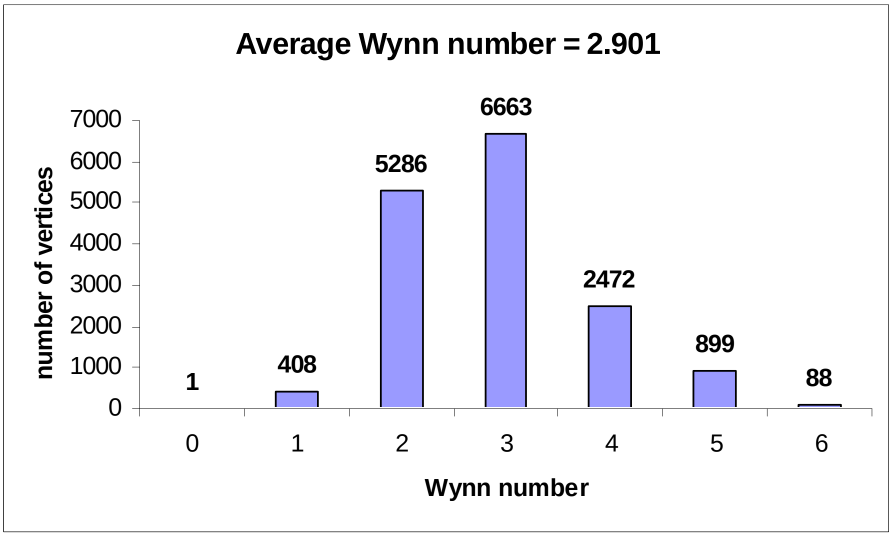

Fig. 2. Number of vertices in the baseball graph with different vaues of Wynn number. Average Wynn number = 2.901

the maximum Bacon numbers of the Hollywood graph are very small, which provides an argument in favor of the "small world hypothesis" mentioned above.

3.3 Baseball Graph and Wynn Number

Collaboration networks similar to the ones mentioned above can also be constructed in sports. One example of such a network is the "baseball graph" representing all baseball players who ever played in the MLB. In this graph, two players are connected if they ever were teammates. The most recent baseball graph has 15817 vertices. Links between any pair of baseball players can be found at the "Oracle of Baseball" website [8].

One can assign the central role in this graph to Early Wynn, a member of the Hall of Fame who spent 23 seasons in the MLB. Figure 2 shows the distribution of Wynn numbers in the baseball graph. The maximum Wynn number is 6, which is smaller than the maximum Bacon number since total number of baseball players is less than the number of Hollywood actors.

3.4 Diameter of Collaboration Networks

Another aspect that should be mentioned here is that the maximum from the central vertex in the collaboration graphs certainly depends on the choice of this central vertex. The reason for choosing Kevin Bacon as the center of the Hollywood graph, and Early Wynn as the center of the baseball graph is the fact that it is reasonable to expect them to be connected to many vertices: Bacon appeared in many movies, and Wynn played in several baseball teams had a lot of teammates during his long career. However, one can choose less "connected" centers of these graphs, and in this case the maximum distance from the new center of the graph may significantly increase. For example, if one chooses Barry Bonds as the center of the baseball graph, the maximum Bonds number will be 9 instead of 6. Moreover, in the Hollywood graph, it is possible to choose the center so that the maximum distance from it is equal to 14, and the average distance is greater than 6 (instead of 2.946). Therefore, in order to have a more complete information about the structure of these graphs, one should calculate the maximum possible distance among *all* pairs of vertices in the graph. Recall that this quantity is referred to as the *diameter* of the graph. Clearly, the diameter can be found by considering each vertex as the center of the graph, calculating corresponding maximal distances, and then choosing the maximum among them.

In the next section, we study the properties of the NBA graph incorporating basketball players playing in the world's best basketball league. In a similar fashion, we introduce the *Jordan number*, investigate its values corresponding to different vertices, and calculate the diameter of this graph.

4 NBA Graph

The NBA graph considered in this section is constructed using the same idea as the graphs described above. Here we provide a detailed description of the structural properties of this graph. As we will see, its properties are rather similar to the properties of other social networks, which confirms the small-world hypothesis.

4.1 General Properties of the NBA Graph

The instance of the NBA graph that we consider in this section is relatively small and contains only those players who are *currently* playing in the NBA (as of the season of 2002-2003). However, this information is sufficient to reveal that the NBA graph follows similar patterns as other social networks. As of May 2003, the total number of players in the rosters of all the NBA teams is equal to 404 (players picked in the 2003 NBA draft and transfers that occurred after the end of the 2002-2003 season are not taken into account). An edge connects two given players if they ever played in the same team. Consequently, the constructed NBA graph has 404 vertices, and 5492 edges connecting them. Note that the maximum possible number of edges is equal to $404 \times (404 - 1)/2 = 81406$, therefore, the *edge density* of this graph (i.e., the ratio of the number of edges to the maximum possible number of edges) is rather small: $5492/81406 = 6.75\%$.

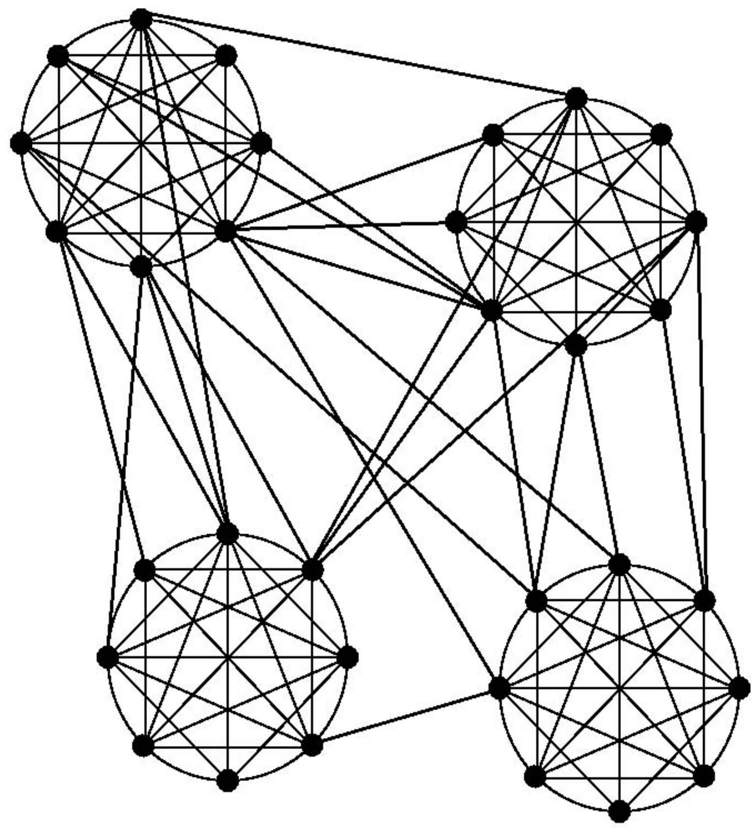

Fig. 3. General structure of the NBA graph and other collaboration networks.

As one can easily see, this graph has a highly specific structure: the players of every team form a *clique* in the graph (i.e., the set of completely interconnected vertices), because all the vertices corresponding to the players of the same team must be interconnected. Since many players change teams during or between the seasons, there are edges connecting the vertices from different cliques (teams). Note that this type of structure is common for all "collaboration networks" (see Figure 3).

It should be pointed out that the number of players in a basketball team is relatively small, and the players' transfers between different teams occur rather often, therefore, it would be logical to expect that the NBA graph should be *connected*, i.e., there is a path from every vertex to every vertex, moreover, the length of this path must be small enough. As we will see below, calculations confirm these assumptions.

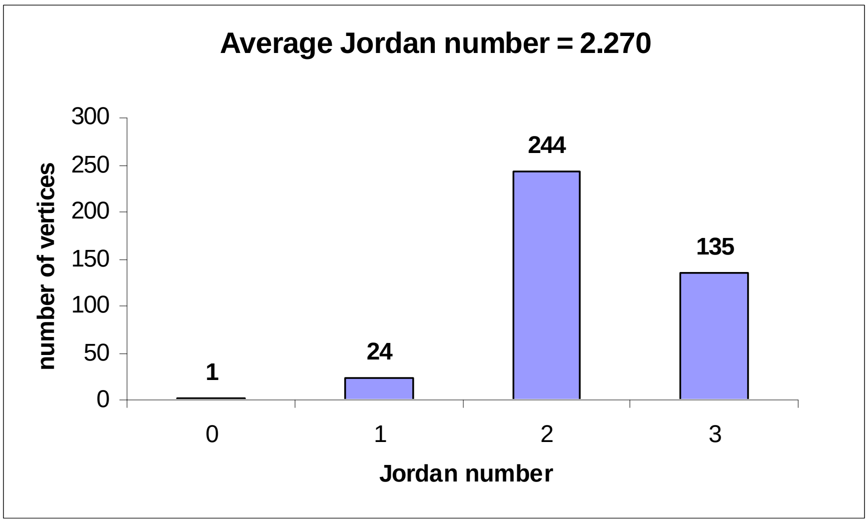

Fig. 4. Number of vertices in the NBA graph with different values of Jordan number. Average Jordan number = 2.270

First, we used a standard breadth-first search technique for checking the connectivity of the considered graph. Starting from an arbitrary vertex, we were able to locate all other vertices in the graph, which means that every vertex is reachable from another, therefore, the graph is connected. In the next subsection, we will also see that every pair of vertices in this graph are connected by a short path, which is in agreement with the "small-world hypothesis".

4.2 Diameter of the NBA Graph and Jordan Number

The next subject of our interest is verifying if the NBA graph follows the small-world hypothesis. We need to answer the question, what is the distance between any two vertices in this graph?

Similarly to the social graphs mentioned above, we define the "central vertex" in the NBA graph corresponding to Michael Jordan, who played for Washington Wizards during his final NBA season. Obviously, all other players in the Wizards' roster for 2002-2003, as well as all the players who have played with Jordan during at least one season in the past, have Jordan number 1. It should be noted that Michael Jordan played only for two teams (Chicago Bulls and Washington Wizards) through his entire career, therefore, one can expect that the number of players with Jordan number 1 is rather small. In fact, only 24 players currently playing in the NBA have Jordan number 1.

Following similar logic, the players who have played with Jordan's "collaborators" have Jordan number 2, and so on. However, it turns out that the

Table 1. Jordan numbers of some NBA stars (end of the 2002-2003 season).

Player	Team	Jordan Number
Kobe Bryant	Los Angeles Lakers	2
Vince Carter	Toronto Raptors	2
Vlade Divac	Sacramento Kings	2
Tim Duncan	San Antonio Spurs	2
Michael Finley	Dallas Mavericks	2
Steve Francis	Houston Rockets	3
Kevin Garnett	Minnesota Timberwolves	3
Pau Gasol	Memphis Grizzlies	3
Richard Hamilton	Detroit Pistons	1
Allen Iverson	Philadelphia 76ers	2
Jason Kidd	New Jersey Nets	2
Toni Kukoc	Milwaukee Bucks	1
Karl Malone	Utah Jazz	2
Stephon Marbury	Phoenix Suns	2
Shawn Marion	Phoenix Suns	2
Kenyon Martin	New Jersey Nets	3
Jamal Mashburn	New Orleans Hornets	2
Tracy McGrady	Orlando Magic	2
Reggie Miller	Indiana Pacers	3
Yao Ming	Houston Rockets	3
Dikembe Mutombo	New Jersey Nets	2
Steve Nash	Dallas Mavericks	2
Dirk Nowitzki	Dallas Mavericks	2
Jermaine O'Neal	Indiana Pacers	2
Shaquille O'Neal	Los Angeles Lakers	2
Gary Payton	Milwaukee Bucks	2
Paul Pierce	Boston Celtics	2
Scottie Pippen	Portland Trail Blazers	1
David Robinson	San Antonio Spurs	2
Arvydas Sabonis	Portland Trail Blazers	2
Jerry Stackhouse	Washington Wizards	1
Predrag Stojakovic	Sacramento Kings	2
Antoine Walker	Boston Celtics	2
Ben Wallace	Detroit Pistons	2
Chris Webber	Sacramento Kings	2

maximum Jordan number in this instance of the NBA graph is only 3, *i.e.*, all the players are linked with Jordan through at most two vertices, which is certainly not surprising: with 29 teams and only around 15 players in each team, NBA is really a "small world". Figure 4 shows the distribution of Jordan numbers in the NBA graph. The average Jordan number is equal to 2.27, which is smaller than the average Bacon number in the Hollywood graph, and the average Wynn number in the baseball graph, due to smaller number of vertices.

Table 1 presents Jordan numbers corresponding to some well-known NBA players. Not surprisingly, most of them have Jordan number 2, except for several players with Jordan number 3: those who joined this league recently, and therefore did not have many teammates through their career, as well as Reggie Miller who spent 16 seasons in the same team (Indiana Pacers), and Kevin Garnett who played in Minnesota for 8 years. Scottie Pippen, Toni Kukoc, and Jerry Stackhouse were Jordan's teammates at different times, therefore, they have Jordan number 1.

Furthermore, we calculated the *diameter* of the NBA graph, i.e., the maximum possible distance between any two vertices in the graph. Since the maximum Jordan number in the NBA graph is equal to 3, one would expect that the value of the diameter to be of the same order of magnitude. As it was mentioned in the previous section, the diameter of the NBA graph can be found as follows: for every given vertex, we calculate the distances between this vertex and all others. In this approach, we need to repeat this procedure 404 times, and every time a different vertex is considered to be the "center" of the graph. Our calculations show that the diameter of the NBA graph (the maximum distance between all pairs of vertices) is equal to 4. Therefore, one can claim that the NBA graph actually follows the small-world hypothesis, since its diameter is small enough.

4.3 Degrees and "Connectedness" of the Vertices in the NBA Graph

As it was pointed out above, the maximum and the average distance from the center of the graph actually depend on the choice of this center. One can easily guess that Michael Jordan is not the most "connected" central vertex of the NBA graph, since he played only for two teams and the number of his former teammates among currently active players is rather small. In fact, the *degree* of the vertex (i.e., the number of edged starting from it, or, the number of teammates) corresponding to Jordan is only 24. Table 2 presents the number of vertices in the NBA graph corresponding to different intervals of the degree values.

It would be reasonable to assume that if one picks a vertex with a high degree as the center of the NBA graph, the average distance in the graph corresponding to this vertex would be smaller than the average Jordan number. We have found the most "connected" players in the NBA graph with the smallest corresponding average distances. Table 3 presents five players who could be the most "connected" centers of the NBA graph. As one can notice, all of them are "bench players" who have changed many teams during their career, therefore, they have high degrees in the NBA graph. Also, an interesting observation is that although Corie Blount's vertex is degree smaller than Jim Jackson's, the average connectivity is higher for Corie Blount, which could be explained by the fact that his teammates were highly "connected" themselves.

Table 2. Degrees of the Vertices in the NBA graph

degree interval	number of vertices
11-20	134
21-30	116
31-40	103
41-50	42
51-60	8
61+	2

Table 3. The most "connected" players in the NBA graph

Player	Team	Degree	Av. Distance
Corie Blount	Chicago Bulls	63	1.906
Jim Jackson	Sacramento Kings	68	1.923
Robert Pack	New Orleans Hornets	57	1.936
Grant Long	Boston Celtics	50	1.946
Bimbo Coles	Boston Celtics	54	1.958

5 Conclusion

In this chapter, we have discussed different properties of social collaboration networks in sports. As one could predict, the properties of these graphs confirm the small-world hypothesis. Investigating these graphs is also interesting from the perspective of "linking" different players, and in many cases it turns out that two players who never played together can be connected by a short chain.

Although the instance of the NBA graph considered in this chapter contains only currently active basketball players, it can be easily extended to reflect all players in the history of the NBA. Moreover, since a lot of foreign players from different countries and continents have come to the NBA in recent years, one would expect that the graph covering all basketball players playing in major foreign championships is also connected and has a small diameter.

Moreover, besides baseball and basketball, similar networks can be constructed for other team games, such as soccer, American football, *etc.* This approach is universal, and it allows one to obtain non-trivial information from player transfers data.

References

1. J. Abello, P.M. Pardalos, M.G.C. Resende, On maximum clique problems in very large graphs, DIMACS Series, 50, American Mathematical Society, 1999, pp. 119-130.
2. W. Aiello, F. Chung, L. Lu, A random graph model for power law graphs, Experimental Math, 10, 2001, pp. 53-66.

3. V. Boginski, S. Butenko, P.M. Pardalos, Modeling and Optimization in Massive Graphs, In: Novel Approaches to Hard Discrete Optimization (P. M. Pardalos and H. Wolkowicz, eds.), American Mathematical Society, 2003, 17-39.
4. V. Boginski, S. Butenko, P.M. Pardalos, On Structural Properties of the Market Graph, In: Innovations in Financial networks (A. Nagurney, ed.), Edward Elgar Publishers, 2003, to appear.
5. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, Graph structure in the Web, Computer Networks, 33, 2000, pp. 309-320.
6. B. Hayes, Graph Theory in Practice, American Scientist, 88: 9-13 (Part I), 104-109 (Part II), 2000.
7. H. Jeong, B. Tomber, R. Albert, Z.N. Oltvai, A.-L. Barabasi, The large-scale organization of metabolic networks, Nature, 407, 2000, pp. 651-654.
8. Oracle of Baseball. http://www.baseball-reference.com/oracle/ Retrieved September 26, 2003.
9. The Oracle of Bacon at Virginia. http://www.cs.virginia.edu/oracle/ Retrieved September 26, 2003.
10. D. Watts, Small Worlds: The Dynamics of Networks Between Order and Randomness, Princeton University Press, 1999.
11. D. Watts, S. Strogatz, Collective dynamics of 'small-world' networks, Nature, 393, 1998, pp. 440-442.