Analysis of Offense Tactics of Basketball Games Using Link Prediction

Tao Zhang, Gongzhu Hu, Qi Liao
Department of Computer Science
Central Michigan University
Mount Pleasant, Michigan, 48859, USA
(zhang3t, hu1g, liao1q)@cmich.edu

Abstract—Every basketball game has a lot of game records, also called match data. All the data are not only statistical but also logical and spatial. People normally use these kind of data to obtain statistical or summarized information of the games, but few have used these data to analyze the teams' tactics. In this paper, we present an approach to analyze the match data for detecting basketball teams' tactic using link prediction method. The main idea is to create a measure for the team offense tactics based on the Basketball Analysis Graph (BA graph) and use link prediction to extract the information about the cooperation between teammates and offense priority. The information may be used for basketball game strategy assistance.

Keywords: basketball offense tactics, game strategy assistance, social networks, link prediction, Katz measure.

I. Introduction

In modern basketball games, the players on the court play with the tactics which designed by the coach group. There are two kinds of tactics. One is offensive tactic and the other is defensive tactic. In a particular basketball game, the offensive tactics are the most dynamic. Enormous amount of basketball game data are available that mostly provide statistical information of individual players and the teams as a whole. Researchers have used these data trying to obtain information that may be useful for basketball teams to make decisions on their game playing strategies. However, we have seen few published work that analyzed the offense behaviors of the rival team that would be useful for coaches to find out the relationship between different players who cooperate to score. In this paper, we present a method to address this problem — detect the offensive tactics from the game data sets, using graph-based link prediction.

In our work, we treat the offensive tactics as relationships between team players expressed as ball-passing chains along with the final actions and results of the tactics. Such relationships are represented as a network (graph) that emulates general social networks. A graph-based link prediction approach is used to find the *centrality* of such networks that represents the potential between particular teammates who cooperate both directly and indirectly in teams' offensive tactic. Katz measure [6] is used in the link prediction with consideration of both path-lengths and edge-weights in the graph. Examples of experiments is presented to illustrate the proposed method

by finding out a considerable scoring cooperative pair using one game data between two NBA basketball team.

II. Background

In this section, we provide some basics of basketball games and different types of game data that are used in our study.

A. Basketball Tactic Basics

Basketball teams have a variety of defensive strategies e.g. "man to man defense" and "zone defense". Besides the general defensive strategies, some specific defensive solution will be applied to different lineup. For instance, A "trap" is set when two defenders double-team the ball-handler, trying to force a turn-over or a jump-ball situation. The motivation of this paper is to find out whether it is a good defensive choice to double or isolate the particular players.

In contrary to defensive tactics, offensive tactics are more dynamic. Although the offensive tactics can also be categorized based on some features, they are in general a lot hard to classify because some of the characteristics are blurred and can only be approximated.

B. Basketball Game Data

In the paper, most of the data examples come from American National Basketball Association (NBA). Most of traditional data types from a basketball game are numerical about all the statistics of game records, including records of individual players as well as records of the teams. These records include points, 3-pointers, field goals, rebounds, etc.

The traditional game records are good at representing the result of games. There are other types of game records that are better for study and analysis called advanced data. In advanced data, player's shot selections are located by court position, including all the positions the player scored in that game and the positions the player missed. The advanced game data are the ones we used in our study.

III. Related Work

We shall briefly discuss the related work in two parts: data analysis particularly for basketball games, and general link prediction method for network analysis.

A. Data Analysis of Basketball Games

The NBA website has a statistic study (`stats.nba.com`) on advance game records, including all kinds of data about the teams and players, such as field points, rebounds, assists, defense, and so on. The site, however, provides analysis of the data but only with statistical approaches.

There are several studies on analysis of basketball games, such as event mining [3], screen strategy analysis using video data [5], game analysis using neural networks [14], outcome prediction [9], [11], trajectory reconstruction [12], and event detection [13].

The study in [15] about basketball winning factors indicated that detection of rival team's tactic plays an important role in winning a basketball game. The task of our work presented in this paper is to extract rival team's offense tactics. Different from [15], our work creates a "social network" (graph) representing defense tactics and uses link prediction on the network to accomplish the task.

B. Link Prediction Method for Network Analysis

Network mining is a data mining task to find information from large networks (structural data), such as social networks. A network represents a physical or social phenomena with vertexes representing the entities in the phenomena and edges representing the relations between the entities. In out study, the network is represented as a *basketball analysis graph* (BA graph) where vertexes are the players and the edges are the pass of the ball from a player to another.

Network analysis has been a topic of research for many years. A thorough study on networks is given in [10], and [2] presented some of the most important developments in the area of social network analysis with quantitative models and methods. One of the fundamental network analysis problems is *link prediction* that tries to estimate or predict the existence of a link between nodes based on the information in the observed links and vertexes [7], [8]. The prediction of links on a network is in general determined by certain quantitative measure (or score) that may be based on neighborhood of vertexes or based on paths. A classical neighborhood-based measure is *Jaccard coefficient* and a path-based measure is *Katz index* (also called *Katz score*) [6]. These methods assign scores to pairs of network nodes (x, y) as a measure of similarity or proximity between x and y. Katz measure is usually used to calculate centrality in directed networks such as citation networks and the World Wide Web [10]. It considers not only the shortest path but all walks between a pair of nodes which fit our requirement.

IV. LINK PREDICTION ON BA GRAPH

The proposed method consists of two main steps: (1) construct a Basketball Analysis Graph (BA graph) from the offensive chain data; (2) apply link prediction to the graph to find "important" links. These steps are described in this section with algorithms and a simple example.

A. Offensive Chain Records

The solution is a special data format called *offensive chain* to collect every offense possession. Let $P = \{p_1, \cdots, p_m\}$ be a team of m players, $A = \{a_1, \cdots, a_k\}$ be a set of k actions (an action may be a field goal, 3-pointer, turnover, technical fault, field fault 2, field fault 3, add-one 2, add-one 3, etc.) and $R = \{scored, failed\}$ be a set of two possible outcomes. An offensive chain is defined as

$$c = (p_i^n, a, r), \ p_i \in P, \ a \in A, \ r \in R \qquad (1)$$

where p_i^n is a sequence of n players participated in one offense ordered by the order in which the ball was passed. An example of four offensive chains is shown in Fig. 1.
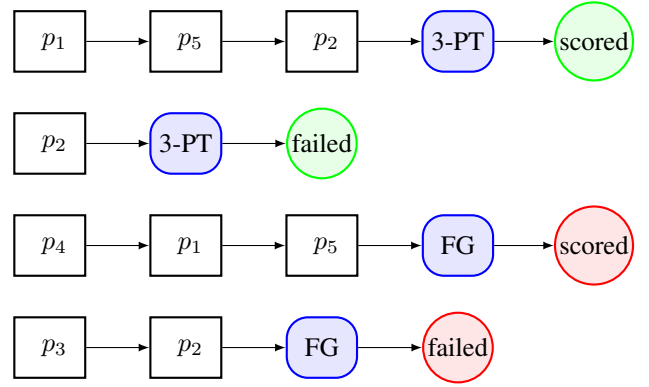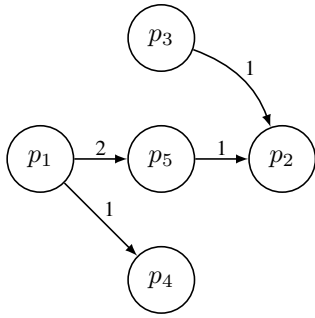

Fig. 1. Offensive chains

Each chain represents an offense that is a sequence of ball-passes from player to player until an action is taken. Each item in the chain is a player who participate the offense except the last two items. The last two items are the types of the action and its result. "3-PT" means a 3 pointer trying to score and "FG" means a field goal trying to score. The end-item of each chain indicates the offense scored or failed. Every team player will be recorded using a unique id. Their positions describe ball's moving track. For example, in the first chain in the figure, player p_1 passed the ball to player p_5 who passed to player p_2, and p_2 made a 3-pointer shot and scored. During one game, there are usually around 100 offensive chains for one team. The specific offensive tactics analysis may study the previous 3 games with the same rival team. The number of offensive chains will rise to around 300.

B. BA Graph

Basketball Analysis Graph (BA graph) is a graph built with offensive chain records as input. The BA graph is a network for which the link prediction is to be applied to extract relations between the players regarding the offense tactics. For example, Fig. 2(a) is a BA graph for the offensive chains given in Fig. 1, and the action and result information are stored in extra data structures associated with the nodes as shown in 2(b). The *weight* on each edge (x, y) is the number of ball-passes from player x to player y

(a) BA graph

Node	Action	Result
p_2	3-PT	scored
	3-PT	failed
p_4	FG	scored
p_5	FG	failed

(b) Action/result information

Fig. 2. BA graph of the offensive chains

Algorithm 1: BuildGraph(C)

Input: $C = \{c_i\}$ — set of offensive chains (Formula (1))
Output: $G = \langle V, E \rangle$ — weighted directed graph

1 **begin**
2      $G \leftarrow \langle \varnothing, \varnothing \rangle$
3   **foreach** $c_i \in C$ **do**
4          $x \leftarrow$ 1st player $p_{i1} \in c_i$
5   **if** $x \notin V$ **then**
6   AddVertex(x, G)
7   **foreach** *next* $p_{ij} \in c_i$ **do**
8              $y \leftarrow p_{ij}$
9   **if** $y \notin V$ **then**
10   AddVertex(y, G)
11              $(x, y, w) \leftarrow (x, y, w + 1)$
12              $x \leftarrow y$
13   **end**
14   **for** *action and result* $a, r \in c_i$ **do**
15   add (a, r) information to $x.table$
16   **end**
17   **end**
18   **return** G
19 **end**

Algorithm 2: AddVertex(v, G)

Input: v — a vertex
Input: $G = \langle V, E \rangle$ — a directed graph
Output: Graph G updated with v, edges (v, u) and (u, v) added for all $u \in V$

1 **begin**
2   **foreach** $u \in V$ **do**
3          $E \leftarrow E \cup \{(v, u, 0)\}$
4          $E \leftarrow E \cup \{(u, v, 0)\}$
5   **end**
6      $V \leftarrow V \cup \{v\}$
7   **return** G
8 **end**

Algorithm 1 outlines the procedure to build the BA graph. It uses Algorithm 2 when a vertex is added to the graph. First, here are some notations used in the algorithm. Let $G = \langle V, E \rangle$ be a BA graph with vertex set V and edge set E. Let $C = \{c_i\}$ be a set of offense chains (defined in Formula (1)). In the vertex set V, each vertex v is associated with action/result/table. Each edge e in the edge set is defined as $e = (v_1, v_2, w)$ with weight w.

Note that we extended the notion of an edge by attaching it with a weight w as denoted by $e = (v_1, v_2, w)$.

Here is a summary of the features of a BA Graph G:

- V is a set of players and their scoring information.
- E is a set of ball-passing time from one player to another.
- $5 \leq |V| \leq$ maximum number of team players. $G(|V| = max)$ is called "team graph." All other graphs are sub-graphs of the "team graph."
- As a complete weighted directed graph, $\forall v \in V$, in-degree(v) = out-degree(v).
- BA graph is often transformed to connected sub-graph(s) with appropriate pruning of low-weighted edges for further analysis.

C. Katz Measure for Link Prediction

There are unlimited plays that can be executed with the different combination of the 5 players on the court. It's sometimes ineffective to detect offense pattern by memorizing and recognizing the ball's moving track. However, for every offensive tactic, there is a starter and a finisher. Finding the most relative offense pair (starter, finisher) can help defender to guard against the right person or the more dangerous player dynamically.

Unlike the traditional analysis, we used link prediction method [4], [7] to find the most closely cooperating pairs of players by considering both directly assisting and indirectly assisting. Especially about the latter, it's difficult to calculate relevance with common statistical analysis. Thus, we applied graph link prediction methods commonly used in social network research to BA graphs. In particular, the Katz index is used for analyzing offensive chains in basketball games.

Let $G = (V, E)$ be the graph representing a network, Katz index is a score of a pair of two vertexes $x, y \in V$:

$$score(x, y) = \sum_{l=1}^{\infty} \beta^l \cdot |paths_{x,y}^{(l)}| \tag{2}$$

where $paths_{x,y}^{(l)}$ is the set of all length-l paths from x and y, and $\beta > 0$ is a parameter (the damping factor) controlling the path weights. It is clear that the terms of large l in the sum contributes much less than the terms of smaller l when β is small. The Katz score can be calculated as unweighted or

weighted [1], [7].

For link prediction, Katz scores can be used as a similarity measure between network nodes to indicate that nodes x and y has a "strong" relation for $score(x, y) > \theta$, where θ is a given threshold parameter. Another reason for choosing Katz score is that the measure will be effected by both direct linked path and other paths which could be more important in finding the most cooperative pair of players while studying players of one team's entire roster. Although Katz score might be affected by directly connected edges of the paths, improper deviation could be inevitable because of the small scale of the BA graph and weighted calculation solution we used. In this research, the number of passings between two players is used as the weight of the BA edges for calculating Katz scores. Unweighted Katz score is not used simply because it may cause improper deviations mentioned above.

We may also calculate the Katz scores for all pairs of players in a complete graph generated as described in Section IV-B. Since there are many edges having low weights (e.g. 0 weight) indicating that the pair of players rarely cooperating in a game, we may prune the graph first so that there will be much less paths to consider when calculating the Katz scores. An edge $e = (u, v)$ with weight w is removed if $w < \alpha$ where α is a selection threshold.

Once the Katz scores are calculated, the pairs of players (x, y) are sorted on the Katz score, and those pairs with $score(x, y) > \theta$ are included in the result of the link prediction. That is, the pairs of players with high Katz scores are considered critical players in the offensive tactics and need to be guarded against by the defense team.

The algorithm of the conditional analyzing for BA graph is outlined in Algorithm 3. Weighted Katz score is calculated according to formula (2).

Algorithm 3: BA-analyzing(G)

Input: BA graph $G = \langle V, E \rangle$ resulted from Algorithm 1
Output: List L of $(x, y, score(x, y)), x, y \in V$

1 **begin**
2    $L \leftarrow \varnothing$
3   Determine the pruning threshold α
4    $G' \leftarrow \text{Prune}(G, \alpha)$
5   **foreach** *pair* $(x, y), x, y \in V$ **do**
6       $L \leftarrow L \cup \{(x, y, score(x, y, G'))\}$
7   **end**
8   Sort(L) on $score$
9   Determine the prediction threshold θ
10   **foreach** $l_i \in L$ **do**
11   **if** $l_i.score < \theta$ **then**
12          $L \leftarrow L - \{l_i\}$
13   **end**
14   **return** L
15 **end**

V. EXPERIMENTS AND EVALUATION

We have applied the proposed method to several offensive chain data sets of basketball games, and the results of one of these data sets are presented in this section. This data set contains 31 offensive chains of LA Lakers in the first quarter of the LA Lakers vs. Portland Trail Blazers game on February 22, 2013. Eight players were involved in the 31 offense actions given in Table I.

TABLE I
OFFENSIVE CHAINS (1ST QUARTER, LAKERS VS. BLAZERS)

No.	Chain	Action	Result
1	$10 \rightarrow 6 \rightarrow 10$	s2	0
2	12	s2	1
3	$10 \rightarrow 6 \rightarrow 10$	to	0
4	24	s2	0
5	$10 \rightarrow 12$	s2	1
6	$10 \rightarrow 24$	s3	0
7	$12 \rightarrow 10$	fn	0
8	$10 \rightarrow 24 \rightarrow 6$	s2	1
9	$10 \rightarrow 24$	s2	1
10	$12 \rightarrow 10$	s2	1
11	$10 \rightarrow 24$	s3	1
12	$15 \rightarrow 10 \rightarrow 6$	s3	1
13	10	s2	0
14	$10 \rightarrow 24$	s2	0
15	10	s2	0
16	$24 \rightarrow 12$	s2f1	1
17	$12 \rightarrow 10 \rightarrow 6$	s3	0
18	$12 \rightarrow 24 \rightarrow 12$	s2	0
19	$15 \rightarrow 10 \rightarrow 15 \rightarrow 4 \rightarrow 24 \rightarrow 4$	s3	1
20	$10 \rightarrow 4$	s2f1	1
21	$12 \rightarrow 10$	to	0
22	10	to	0
23	$24 \rightarrow 20 \rightarrow 24$	s2	1
24	$10 \rightarrow 12$	s2	0
25	$12 \rightarrow 10 \rightarrow 20$	s2	0
26	$4 \rightarrow 10 \rightarrow 24$	s2	1
27	$12 \rightarrow 10 \rightarrow 4 \rightarrow 20$	s3	0
28	$12 \rightarrow 20$	s2f2	1
29	$6 \rightarrow 5 \rightarrow 24$	s2	0
30	24	to	0
31	24	s2	1

The results 0 and 1 mean *scored* and *failed*, respectively. The actions in the table are:

s2:   field goal for 2
s3:   pointer
to: turnover
fn: foul but no shot of foul
s2f1:   field goal for 2, foul shot for 1/free throw 1 of 2
s2f2:   field goal for 2, foul shot for 2/free throw 2 of 2

The corresponding BA graph, after edges with $w = 0$ are removed, is shown in Fig. 3. In the figure, a bi-directional edge between two nodes $(x, y), x < y$, has two weights w_1 and w_2 representing $(x \xrightarrow{w_1} y)$ and $(y \xrightarrow{w_2} x)$.

From the BA graph, both unweighted and weighted Katz scores are calculated for nodes $v_i, v_j \in V$, given in Table
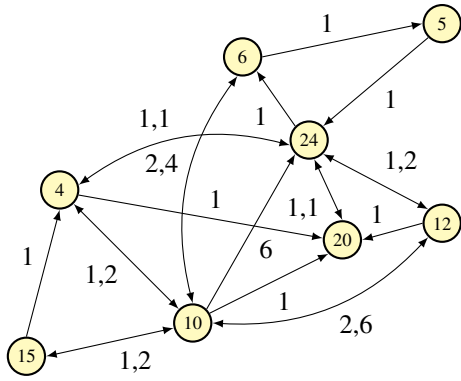
Fig. 3. BA graph after pruning of edges with $w = 0$. An edge $x \xleftarrow{w_1,w_2} y$ with two weights represents $(x \xrightarrow{w_1} y)$ and $(y \xrightarrow{w_2} x)$ for $x < y$. That is, the first weight is from smaller-numbered vertex to larger-numbered vertex, and the second weight is from larger-numbered vertex to smaller-numbered vertex.

II, sorted from the highest to lowest on the weighted scores. $\beta = 0.5$ was used in the calculation.

TABLE II
KATZ SCORES

| Pair No.	v_i	v_j	Katz Score	
			Unweighted	Weighted
1	12	20	0.677	9.766
2	10	20	0.670	9.344
3	12	24	0.827	8.375
4	10	24	0.895	7.375
5	10	6	0.749	6.156
6	24	20	0.222	6.031
7	4	20	0.400	6.000
8	10	12	0.827	5.500
9	10	4	0.774	4.875
10	12	10	0.608	4.438
11	4	24	0.876	4.375
12	15	4	0.492	3.813
13	24	4	0.407	3.313
14	4	10	0.608	3.063
15	15	10	0.675	2.969
16	24	6	0.586	2.750
17	24	12	0.758	2.125
18	6	10	0.656	1.875
19	5	24	0.500	0.500
20	10	15	0.500	0.500
21	20	24	0.500	0.500
22	6	5	0.500	0.500

The results show that the unweighted and weighted scores are very different in terms of the relative strength of the player pairs. For example, (12, 24) has high Katz scores for both weighted and unweighted measures, whereas (4, 24) has high weighted measure but low unweighted measure.

Fig. 4 shows the distribution plot of the Katz scores, with weighted in 4(a) and unweighted in 4(b). We also applied this procedure to a subgraph consists of only the offensive chains that were successful (scored), shown in Fig. 5. The weighted Katz scores on this graph is given in Table III, where we use the same Pair No. as in Table II. The score distribution is

shown in Fig. 6.
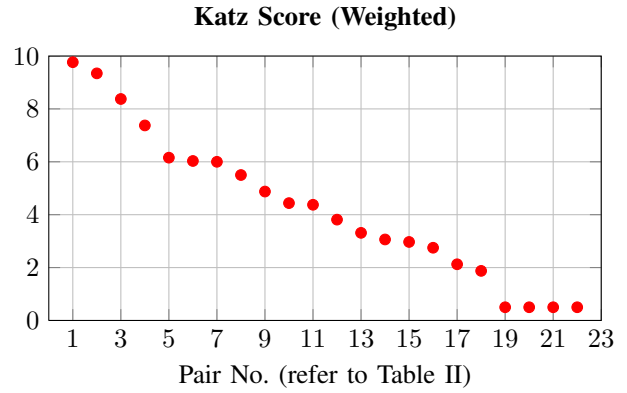
(a) Weighted Katz scores
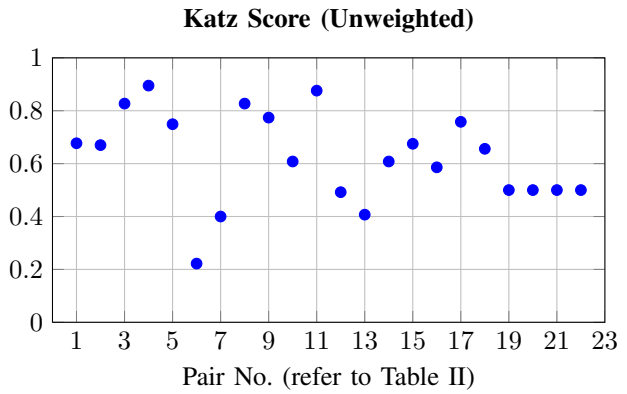
(b) Unweighted Katz scores

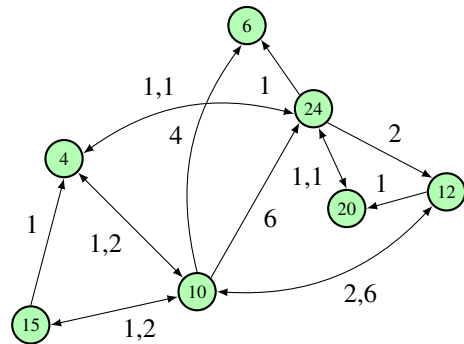Fig. 4. Distribution of the Katz scores

Fig. 5. BA graph of succeeded offensive chains. See caption of Fig. 3 for bi-directional edges.

The results in Tables II and III show that the pairs (10, 24) and (10, 6) (Pairs No. 4 and 5) had high Katz measures in unweighted, weighted, as well as weighted–succeeded cases. This result is consistent with the fact that player 10 (Steve Nash) had the highest *assists*, player 24 (Kobe Bryant) had the highest *points* and player 6 (Earl Clark) had the second highest *points*, as shown in the official data in Table IV. The link analysis method revealed information about the collaborations

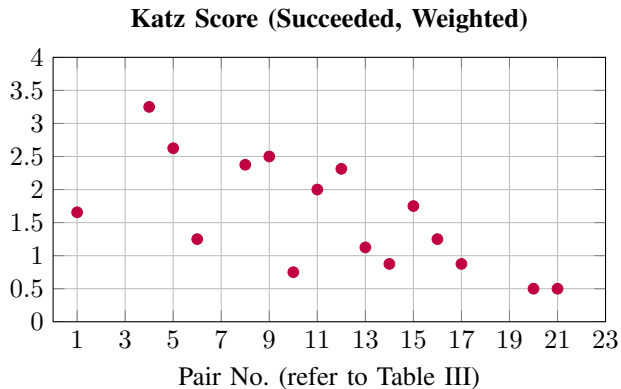| Pair No.	v_i	v_j	Weighted Katz Score
1	12	20	1.656
4	10	24	3.250
5	10	6	2.625
6	24	20	1.250
8	10	12	2.375
9	10	4	2.500
10	12	10	0.750
11	4	24	2.000
12	15	4	2.313
13	24	4	1.125
14	4	10	0.875
15	15	10	1.750
16	24	6	1.250
17	24	12	0.875
20	10	15	0.500
21	20	24	0.500

Fig. 6. Distribution of the Katz scores of succeeded pairs

between team players that characterize the game playing tactics that are not easy to find in regular statistical data.

TABLE IV
1ST COURT LAL DATA RECORD BY OFFICIAL PLAYBOOK

No PLAYER		MIN	FG	FGA	3P	3PA	FT	FTA	OR	DR	TOT	A	PF	ST	TO	BS	PTS
15 Metta World Peace	F	7:27	0	0	0	0	0	0	0	2	2	0	1	0	0	0	0
6 Earl Clark	F	8:00	2	3	1	2	0	0	0	1	1	0	1	0	0	0	**5**
12 Dwight Howard	C	10:32	2	4	0	0	1	2	1	4	5	0	1	0	0	0	5
24 Kobe Bryant	G	12:00	5	9	1	2	0	0	1	1	2	2	0	1	1	0	**11**
10 Steve Nash	G	10:32	1	4	0	0	0	1	1	2	**4**	0	1	3	0	2
4 Antawn Jamison		5:28	1	2	1	1	2	2	0	0	0	0	0	1	0	0	5
20 Jodie Meeks		4:33	0	1	0	1	2	2	0	0	0	1	0	0	0	0	2
5 Steve Blake		1:28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a graph structure (network) for basketball data and analyzed the the offense behaviors with a graph-based link prediction method.

Our method is focused on the cooperating relationships between players rather than on individual players.

The novel idea of this method is to represent the offensive chains as a weighted directed graph, and use link prediction approach to analyze the graph like a social network. Katz measure was used in the link prediction analysis with the Katz scores calculated not only on the lengths of the paths in the graph, but also on the weights of the edges. Experiments were conducted that showed promising results of the method.

We are currently working on several problems on this topic. First, Data collection and preparation is one of the most challenging tasks of our study. In addition to existing data, we may have to extract game tactics from broadcasting videos or using special recording solution like STATS system. Organize data in a multi-dimensional model with dimensions like team, lineup, roster, and time, so data from different views can be used in analysis for different tasks.

Different measures beside Katz measure will be studied regarding their appropriateness to the applications in network analysis and basketball. Certainly, more experiments are being conducted to better evaluate the proposed method.

REFERENCES

[1] Evrim Acar, Daniel M. Dunlavy, and Tamara G. Kolda. Link prediction on evolving data using matrix and tensor factorizations. In *IEEE International Confernece on Data Mining Workshop*, pages 262–269, 2009.

[2] Peter J. Carrington, John Scott, and Stanley Wasserman. *Models and Methods in Social Network Analysis*. Cambridge University Press, 2005.

[3] Yung-Hui Chen and Lawrence Y. Deng. Event mining and indexing in basketball video. In *Fifth International Conference on Genetic and Evolutionary Computing*, pages 247–251, 2011.

[4] Michael Fire, Lena Tenenboim, Ofrit Lesser, Rami Puzis, Lior Rokach, and Yuval Elovici. Link prediction in social networks using computationally efficient topological features. In *IEEE Third International Confernece on Social Computing*, pages 73–80, 2011.

[5] Tsung-Sheng Fu, Hua-Tsung Chen, Chien-Li Chou, Wen-Jiin Tsai, and Suh-Yin Lee. Screen-strategy analysis in broadcast basketball video using player tracking. In *IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4, 2011.

[6] Leo Katz. A new status index derived from sociometric index. *Psychometrika*, 18(1):39–43, 1953.

[7] David Liben-Nowell and Jon Kleinberg. The link prediction problem for social networks. *Journal of the American Society for Information Science Technology*, 58(7):10191031, 2007.

[8] Linyuan Lü and Tao Zhou. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150–1170, 2011.

[9] Dejan Miljković, L. Gajić, Aleksandar Kovačević, and Zora Konjović. The use of data mining for basketball matches outcomes prediction. In *8th International Symposium on Intelligent Systems and Informatics (SISY)*, pages 309–312, 2010.

[10] Mark E. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[11] Krzysztof Trawiński. A fuzzy classification system for prediction of the results of the basketball games. In *IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 1–7, 2010.

[12] Lifang Wu, Wei Shen, Xiuli Han, Xianglong Meng, and Chao Liu. A method of the basketball trajectory reconstruction. In *9th International Conference on Signal Processing*, pages 1360–1363, 2008.

[13] Min Xu, Ling-Yu Duan, Changsheng Xu, M. Kankanhalli, and Qi Tian. Event detection in basketball video using multiple modalities. In *Proceedings of Fourth International Conference on Information, Communications and Signal Processing*, volume 3, pages 1526–1530, 2003.

[14] Ivanković Zdravko, Racković M., Branko Markoski, D. Radosav, and Miodrag Ivković. Analysis of basketball games using neural networks. In *11th International Symposium on Computational Intelligence and Informatics*, pages 251–256, 2010.

[15] Xianjiang Zhou, Zhi Gao, and Qiaoling Zhang. Analysis approach of winning factors in competitive basketball. In *IEEE International Symposium on Knowledge Acquisition and Modeling Workshop*, pages 1141–1144, 2008.