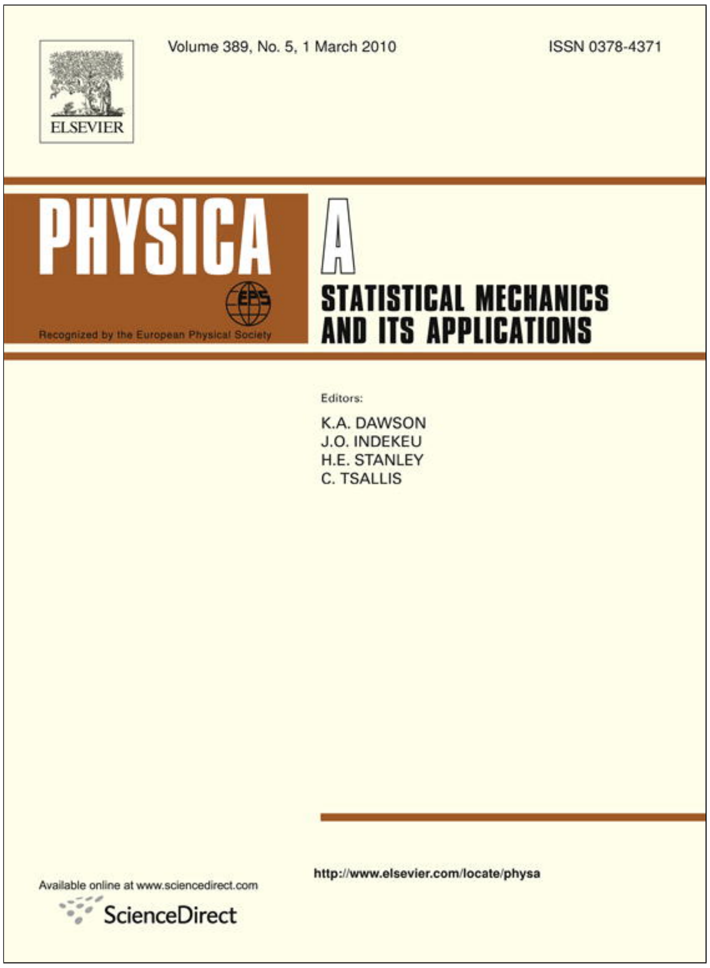

Volume 389, No. 5, 1 March 2010
ISSN 0378-4371

ELSEVIER

PHYSICA A

STATISTICAL MECHANICS
AND ITS APPLICATIONS

Recognized by the European Physical Society

http://www.elsevier.com/locate/physa

Available online at www.sciencedirect.com
ScienceDirect

Physica A 389 (2010) 1131–1141

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Mutually-antagonistic interactions in baseball networks

Serguei Saavedra [a,*], Scott Powers [b], Trent McCotter [c], Mason A. Porter [d,e], Peter J. Mucha [b,f]

[a] *Kellogg School of Management and NICO, Northwestern University, Evanston, IL, 60208, USA*
[b] *Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599-3250, USA*
[c] *School of Law, University of North Carolina, Chapel Hill, NC 27599-3380, USA*
[d] *Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford, OX1 3LB, UK*
[e] *CABDyN Complexity Centre, University of Oxford, Oxford, OX1 1HP, UK*
[f] *Institute for Advanced Materials, Nanoscience and Technology, University of North Carolina, Chapel Hill, NC 27599-3216, USA*

ARTICLE INFO

Article history:
Received 7 August 2009
Received in revised form 25 October 2009
Available online 5 November 2009

PACS:
64.60.aq
02.50.-r
05.40.Fb
87.23.-n

Keywords:
Bipartite networks
Ranking systems
Random walkers
Competition dynamics

ABSTRACT

We formulate the head-to-head matchups between Major League Baseball pitchers and batters from 1954 to 2008 as a bipartite network of mutually-antagonistic interactions. We consider both the full network and single-season networks, which exhibit structural changes over time. We find interesting structure in the networks and examine their sensitivity to baseball's rule changes. We then study a biased random walk on the matchup networks as a simple and transparent way to (1) compare the performance of players who competed under different conditions and (2) include information about which particular players a given player has faced. We find that a player's position in the network does not correlate with his placement in the random walker ranking. However, network position *does* have a substantial effect on the robustness of ranking placement to changes in head-to-head matchups.

1. Introduction

The study of networks has experienced enormous growth in recent years, providing foundational insights into numerous complex systems ranging from protein interaction networks in biology to online friendship networks in the social sciences [1–3]. Research on ecological and organizational networks has provided a general framework to study the mechanisms that mediate cooperation and competition dynamics between individuals [4–9]. In such networks, competitive interactions result from the indirect competition between members of different populations who either compete for the same resources or are linked through consumer–resource relationships. However, data on mutually-antagonistic interactions—which occur between individuals who directly fight or compete against each other—have been more difficult to collect [10,11]. Mutually-antagonistic interactions also occur frequently in different social contexts, such as sports. In the present paper, we consider head-to-head matchups between Major League Baseball (MLB) pitchers and batters: Pitchers benefit by "defeating" batters and vice versa. Using data from www.retrosheet.org [12], we characterize the more than eight million MLB plate appearances from 1954 to 2008. We consider full careers by examining head-to-head matchups over a multi-season ("career") network and single-season performances by constructing networks for individual seasons.

Major League Baseball uses votes by professional journalists to recognize career achievement of players through induction into a Hall of Fame (HOF) and single-season performance through awards such as Most Valuable Player (MVP) and

---

[*] Corresponding author.
E-mail address: s-saavedra@northwestern.edu (S. Saavedra).

0378-4371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2009.10.038

Cy Young (for pitching performance) [13]. Although the HOF purports to recognize the best players of all time, the selection of players to it is widely criticized by fans and pundits each year because of the lack of consistency when, e.g., comparing players from different eras, who play under fundamentally different conditions—in different ballparks, facing different players, etc. [14–16]. Such arguments come to the fore when attempting to draw comparisons between players elected to the HOF and others who did not make it. For instance, how can one tell whether Jim Rice (elected to the HOF in 2009) had a better career than Albert Belle (who dropped off the ballot because of low vote totals after only two years [17])? Does Bert Blyleven, who appeared on 62.7% of the HOF ballots in 2009 – short of the 75% required for election – belong in the HOF? Is Sandy Koufax, who played from 1955–1966 and is in the HOF, better than Pedro Martinez (an active player), who will presumably eventually be elected to the HOF? To address such questions, it is insufficient to rely purely on raw statistics; one must also consider quantitative mechanisms for comparison between athletes who played under different conditions. We take a first, simple step in this direction through the study of biased random walkers [18,19] on baseball networks. This allows us not only to construct a quantitative, systematic, and transparent ranking methodology across different eras but also to investigate the interplay between these dynamics and the underlying graph structure. We thereby reveal key properties of mutually-antagonistic interactions that can potentially also be applied in other settings.

While "water-cooler" discussions about which players should and should not be enshrined in the HOF (and, more generally, how to rank players) can often be fascinating, as indicated by the above paragraph, the primary goal of our paper is to investigate interesting features of the baseball *networks* and the impact that network structure can have on rankings. It is necessary to include some example rank orderings to have a proper context for such a discussion, but it is important to note that the rankings we show in the present paper must be taken with several grains of salt because our efforts at simplicity, which are crucial to highlighting the interplay between network structure and player rankings, require us to ignore essential contributing factors (some of which we will briefly discuss) that are necessary for any serious ranking of baseball players.

The rest of this paper is organized as follows. In Section 2, we define and characterize the mutually-antagonistic baseball networks and study the time evolution of various graph properties. In Section 3, we provide a description of the biased random walker dynamics that we employ as a ranking methodology across eras and for single-season networks. In Section 4, we study the interplay between the random walker dynamics and graph structure, paying special attention to the sensitivity of the player rankings. In Section 5, we conclude the paper and discuss a number of potential applications of our work. We explain additional technical details in two appendices.

2. Network characterization and evolution

We analyze baseball's mutually-antagonistic ecology by considering bipartite (two-mode) networks of head-to-head matchups between pitchers and batters. As shown in Fig. 1, bipartite networks are formed using two disjoint sets of vertices, P (pitchers) and B (batters), and the requirement that every edge connect a vertex in P to one in B [8,20,21]. (Note that we track the pitching and batting performances of pitchers as two separate nodes.) We consider such interactions in terms of three different bipartite representations (with corresponding matrices): (1) The binary matchups $\mathbf{A}$ in which the element A_{ij} equals 1 if pitcher i faced batter j at any point and 0 otherwise; (2) the weighted matchups $\mathbf{W}$ in which the element W_{ij} equals the number of times that i faced j; and (3) the weighted outcomes $\mathbf{M}$ in which the element M_{ij} equals a "score" or performance index, which is determined using what are known in baseball as "sabermetric" statistics (see Section 3) [14,15,22] and which characterizes the results of all matchups between i and j. For each of these bipartite pitcher–batter networks, we also utilize corresponding square adjacency matrices:

$$\hat{\mathbf{A}} = \begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{pmatrix}, \qquad \hat{\mathbf{W}} = \begin{pmatrix} \mathbf{0} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{0} \end{pmatrix}, \qquad \hat{\mathbf{M}} = \begin{pmatrix} \mathbf{0} & -\mathbf{M} \\ \mathbf{M}^T & \mathbf{0} \end{pmatrix},$$

so that they are appropriately symmetric ($\hat{\mathbf{A}}$ and $\hat{\mathbf{W}}$) and anti-symmetric ($\hat{\mathbf{M}}$). We construct and analyze each of these representations for the single-season networks and the aggregate (career) network that contains all pitcher–batter interactions between 1954 and 2008.

To identify the changes in time in the organization of baseball networks, we examine the graph properties of single-season networks. The number of distinct opponents per player, given by the distribution of player degree $k_i = \sum_j \hat{A}_{ij}$, follows an exponential distribution for a large range and then has an even faster decay in the tail (see Fig. 2). (A recent study has observed power-law behavior for other cumulative quantities in baseball networks using different data sets and observation periods [23].) The mean values of the geodesic path length between nodes and of the bipartite clustering coefficient are only somewhat larger than what would be generated by random assemblages (see Appendix A). However, as with mutually-beneficial interactions in ecological networks [24], the mutually-antagonistic baseball matchup networks exhibit nontrivial relationships between player degree and player strength $s_i = \sum_j \hat{W}_{ij}$, which represents the total number of opponents of a player (counting multiplicity) [1,20]. As shown in Fig. 3A, the relation between strength and degree is closely approximated by a power-law $s \sim k^{\alpha}$ that starts in 1954 at $\alpha \approx 1.64$ for pitchers and $\alpha \approx 1.41$ for batters but approaches $\alpha \approx 1$ for each by 2008. The 54-year trend of a decreasing power-law exponent indicates that real-life events such as the increase in the number of baseball teams through league expansion (e.g., in the 1960s, 1977, 1993, and 1998), the reorganization of the leagues (e.g., in 1994, to three divisions in each league instead of two), and the introduction of interleague play (in 1997) and unbalanced schedules (in 2001)

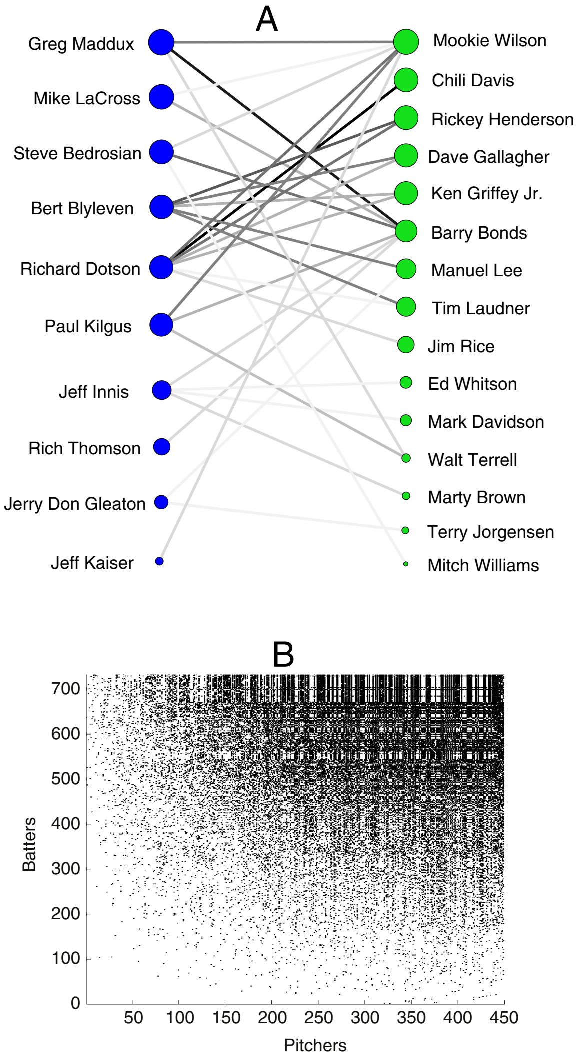

Fig. 1. Bipartite baseball networks. (A) A subset of the bipartite interactions between pitchers (left column) and batters (right column) during the 1989 baseball season. The area of each circle is determined by the node degree (i.e., how many different opponents were faced). Each line indicates that a given pitcher faced a given batter, and the darkness of each line is proportional to the number of plate appearances that occurred (i.e., the node strength). (B) The matrix encoding the complete set of bipartite interactions from 1989, with pitchers (columns) and batters (rows) arranged from the lowest to the highest node degree. An element of the matrix is black if that particular pitcher and batter faced each other and white if they did not. Observe the presence of a core of high-degree players that are heavily connected to each other (top right corner), an important presence of asymmetric interactions (i.e., high-degree players connected to low-degree players), and a dearth of connections between low-degree players (bottom left corner), which are all characteristics of nested networks [21]. Some of the batters are actually pitchers (e.g., Mitch Williams), as National League pitchers (and, since 1997, American League pitchers as well) have a chance to bat and face a small number of pitchers while at the plate.

have modified the organizational and team-competition properties of the networks (also see relevant discussions in Refs. [23,25]). Fascinatingly, this long-term decreasing trend in power-law exponent also seems to exhibit non-punctuated behavior that does not have an obvious explanation via known changes in baseball scheduling or rules. Understanding the origins of this observation thus remains an interesting open question.

An important property mediating the competition dynamics of mutualistic networks in ecology is *nestedness* [9]. Although the definition of nestedness can vary, a network is said to be nested when low-degree nodes interact with proper subsets of the neighbors of high-degree nodes [21] (see Fig. 1). To calculate the aggregate nestedness in the binary matchup network **A**, we employed the nestedness metric based on overlap and decreasing fill (NODF) [26], which takes values in the interval [0, 1], where 1 designates a perfectly-nested network (see Appendix A). Fig. 3B (black circles) shows that single-season baseball networks consistently have nestedness values of approximately 0.28. This value is slightly but consistently higher than those in randomized versions of the networks with similar distribution of interactions (red squares) [21], which we also observe to decrease slightly in time. In common with bipartite cooperative networks [8], this confirms that nestedness is a significant feature of these mutually-antagonistic networks.

Although nestedness is defined as a global characteristic of the network, we can also calculate the individual contribution of each node to the aggregate nestedness [26]. Comparing node degrees and individual nestedness (see Appendix A) before 1973, batters and pitchers collapse onto separate curves (see Fig. 3C). Starting in 1973, however, each of these groups of nodes splits into two curves (see Fig. 3D), corresponding to players in the two different leagues: the American League (AL)

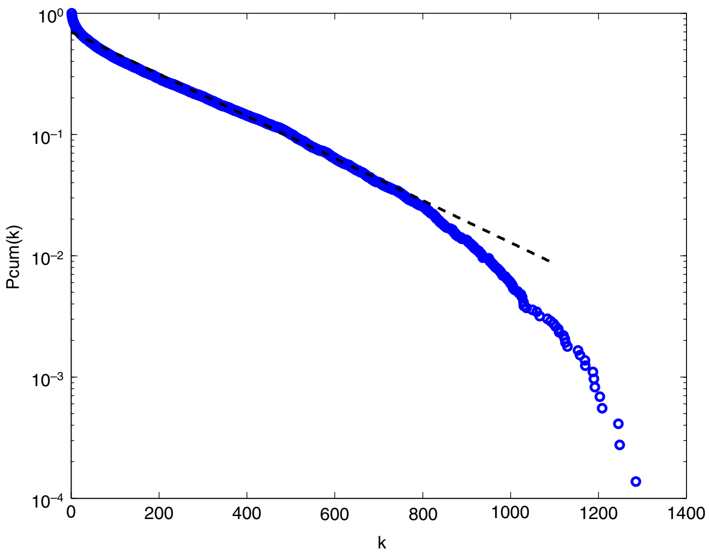

Fig. 2. (Color online) Cumulative degree distribution. Semi-log plot of the cumulative degree distribution $P_{cum}(k)$ for pitchers and batters in the career (1954–2008) network. The dashed line correspond to a theoretical exponential distribution.

and the National League (NL). This structural change presumably resulted from the AL's 1973 introduction of the designated hitter (DH), a batter who never fields but bats in place of the team's pitchers (see Fig. 1), apparently causing the AL to become less nested due to the replacement of low-degree batting pitchers with higher-degree DHs. As we discuss below, this suggests that the network position of a player might affect his own ranking (while, of course, network position is strongly influenced by a player's longevity and thus by his performance). We examined ecological mutualistic networks with similar aggregate nestedness values but found no correlation between degree and individual nestedness in those examples.

3. Biased random walkers

To compare the performance of players, we rank them by analyzing biased random walkers on the bipartite network **M** encoding the outcomes of all mutually-antagonistic interactions between each player pair. Our method generalizes the technique we previously used for NCAA football teams [18,19], allowing us to rank players in individual seasons and in the 1954–2008 career network. It thereby yields a quantitative, conceptually-clear method for ranking baseball players that takes a rather different approach from existing sabermetric methods used to project player performance. Such existing methods include DiamondMind (which uses Monte Carlo simulations), PECOTA (which uses historical players as a benchmark), and CHONE (which uses regression models) [27,28].

To describe the aggregate interaction M_{ij} between pitcher i and batter j, we need to quantify each possible individual pitcher–batter outcome. For simplicity, we focus on the quantity *runs to end of inning* (RUE) [15], which assigns a value to each possible plate appearance outcome (single, home run, strikeout, etc.) based on the expected number of runs that a team would obtain before the end of that inning, independent of the situational context (see Appendix B for specific values). Higher numbers indicate larger degrees of success for the batter. For each season, we add the RUE from each plate appearance of pitcher i versus batter j to obtain a cumulative RUE for the pair. Note that any performance index that assigns a value to a specific mutually-antagonistic interaction can be used in place of RUE without changing the rest of our ranking algorithm. We define the single-season outcome element M_{ij} by the cumulative extent to which the batter's outcome is better ($M_{ij} > 0$) or worse ($M_{ij} < 0$) than the mean outcome over all pitcher–batter matchups that season. When defining the career outcome element M_{ij} for 1954–2008, we account for the offensive inflation in baseball's modern era [14,15] by summing over individual seasons (i.e., we examine outcomes relative to mean outcomes on a per season basis).

We initiate our ranking methodology by considering independent random walkers who each cast a single vote for the player that it believes is the best. Each walker occasionally changes its vote with a probability determined by considering the aggregate outcome of a single pitcher–batter pairing, selected randomly from those involving its favorite player, and by a parameter that quantifies the bias of the walker to select the winner of the accumulated outcome. A random walker that is considering the outcome described by this matchup is biased towards but not required to choose the pitcher (batter) as the better player if $M_{ij} < 0$ ($M_{ij} > 0$).

The expected rate of change of the number of votes cast for each player in the random walk is quantified by a homogeneous system of linear differential equations $\mathbf{v}' = \mathbf{D} \cdot \mathbf{v}$, where

$$D_{ij} = \begin{cases} \hat{W}_{ij} + r\hat{M}_{ij}, & i \neq j \\ -s_i + r\sum_k \hat{M}_{ik}, & i = j. \end{cases} \tag{1}$$

The long-time average fraction of walkers $\tilde{v}_j$ residing at (i.e., voting for) player j is then found by solving the linear algebraic system $\mathbf{D} \cdot \tilde{\mathbf{v}} = \mathbf{0}$, subject to an additional constraint that $\sum_j \tilde{v}_j = 1$. If the bias parameter $r > 0$, then successful players will on average achieve higher rankings than unsuccessful players. For $r < 0$, the random walker votes will instead tend toward the "loser" of individual matchups.

Eq. (1) gives a general one-parameter system for a biased walker with probabilities that are linear in RUE, but the approach is easily generalized by using other functional forms to map observed plate appearance outcomes (in $\mathbf{M}$) into selection probabilities. By restricting our attention to a form that is linear in RUE, the interpretation that the off-diagonal components of $\mathbf{D}$ correspond to random walker rate coefficients requires that these components remain non-negative, a preferable state that leads to a number of beneficial properties in the resulting matrix. For example, this allows us to apply the Perron–Frobenius theorem, which guarantees the existence of an equilibrium $\tilde{\mathbf{v}}$ with strictly positive entries (and similarly guarantees the existence of positive solutions in algorithms such as PageRank) [19,20,29,30]. In practice, this requirement is equivalent in the baseball networks to $|r| \lesssim 0.7$, so that the result of a home run in a single plate appearance matchup (i.e., the case in which a batter faces a pitcher exactly once and hits a home run in that appearance) maintains a small but non-negative chance that a random walker will still select the pitcher.

Because the aggregate outcome of most pairings remains close to the mean, the bias in the random walk is small, and the rankings become essentially independent of the bias parameter. The linear expansion in bias r thereby yields a ranking with no remaining parameters beyond the statistically-selected RUE values. This expansion is $\tilde{\mathbf{v}} = \mathbf{v}^{(0)} + r\mathbf{V} + O(r^2)$. Generalizing the similar expansion described in detail in Ref. [19], the zeroth-order term results in a constant number of votes per player. The additional contribution at first order is given by the solution of a discrete Poisson equation on the graph:

$$\sum_j L_{ij} V_j = \frac{4}{n} \sum_j \hat{M}_{ij}, \tag{2}$$

subject to the neutral charge constraint $\sum_j V_j = 0$. (By analogy with electrostatics, we refer to V_j as the RUE 'charge' of node j.) In Eq. (2), $n = P + B$ is the total number of players, $\mathbf{L} = \mathbf{S} - \hat{\mathbf{W}}$ is the graph Laplacian, $\mathbf{S}$ is the diagonal matrix with elements $s_{ii} = \sum_j \hat{W}_{ij}$ (and $s_{ij} = 0$ for $i \neq j$). We restrict our attention to the first-order ranking that is specified by $\mathbf{V}$ and obtained using the solution of equation Eq. (2).

We tabulate this rank ordering separately for pitchers and batters for both individual seasons and the career network. We compare the results of the random walker ranking to major baseball awards in Table 1. We note that the rankings are highly correlated with the underlying RUE per plate appearance of each player (the correlation is $\rho \approx .96$ for 2008; we obtain similar values for other seasons), so that the top players in the rankings produced by our method have a strong but imperfect correlation with the lists produced by ranking players according to (suitably normalized) raw RUE values. For instance, Todd Helton, who is widely lauded by sabermetrics gurus as a significantly underrated player even when taking into account the likely inflation of his raw statistics from his home ballpark (Coors Field, which highly-ranked batter Larry Walker also called home for several seasons) [31], ranks third among all batters in the career network in RUE values, which we consider in each year relative to the annual average (i.e., before any network structure is taken into account) [32]. One similarly finds a strong correlation between rankings with and without network information using any other sabermetric quantity that one might consider in place of RUE. That is, although the differences in rank ordering that result from considering a player's position in the network are typically small, they are still present: it matters which players one has faced, and that is codified by the network. For example, the differences between random walker rankings and raw RUE rankings appear to appropriately capture the caliber of opponents (e.g., pitchers from teams with relatively anemic offenses – such as the 2008 Nationals, Astros, and Reds – tend to get a boost in their random walker ranking, reflecting that they never had the good fortune of facing the batters on their own teams). We also compared our rankings with a leading contender in baseball analysis, ESPN's MLB Player Ratings, which combines ratings from ESPN, Elias, Inside Edge, and *The Baseball Encyclopedia* [33]. Of the top 99 players who are listed in the 2008 Player Ratings, 12 did not meet our threshold for plate appearances. Comparing the random walker results for the remaining 87 players with the Player Ratings yields a correlation of $\rho \approx .5601$. We thus proceed to study the random walker results for the career network both with confidence that it correlates with methods that are currently used for single-season analysis and caution that the ranking details do not capture all effects according to current best practices in quantitative baseball analysis [34].

The career ranking allows credible comparisons between players from different eras. Considering the rankings restricted to individuals who played in at least 10 seasons (i.e., HOF-eligible players) during the time capsulated in the employed data set, we find that Barry Bonds (batter), Pedro Martinez (starting pitcher), and Mariano Rivera (relief pitcher) are the best players in their respective categories from 1954 to 2008. We show additional rankings in Table 2. Returning to some of our motivating examples, we note that Albert Belle (29th among batters) is ranked much higher than Jim Rice (115th), suggesting that Belle's hitting performance perhaps merits HOF membership more than that of Rice. Similarly, Bert Blyleven ranks higher not only than current HOF competitors such as Jack Morris and Tommy John but also higher than three HOF pitchers with over 300 wins (Steve Carlton, Phil Niekro, and Don Sutton), which is one traditional benchmark for selecting elite pitchers. Direct comparisons with other rank orderings of players across different eras would necessitate restricting the data to sufficiently similar time periods and is beyond the scope of the present study.

Table 1
Single-season awards and random walker rankings. We show the MVP and CY Young award winners for various years from 1954 to 2008. In parentheses, we give the ranking of the player within his own category (pitcher or batter) that we obtained using our random walker ranking system applied to the corresponding baseball season. For most of the seasons, there is reasonable agreement between award winners and their random walker ranking considering that the ranking only considers batter events, thus ignoring other ways in which players contribute. (Note that the Cy Young award was awarded to a single pitcher—rather than one from each league—until 1967.)

	1954	1958	1963
MVP (AL)	Yogi Berra (11th)	Jackie Jensen (8th)	Elston Howard (20th)
MVP (NL)	Willie Mays (2nd)	Ernie Banks (6th)	Sandy Koufax (1st)
Cy Young (AL)	N/A	Bob Turley (14th)	Sandy Koufax (1st)
Cy Young (NL)	N/A	Bob Turley (14th)	Sandy Koufax (1st)
	1968	1973	1978
MVP (AL)	Denny McLain (4th)	Reggie Jackson (11th)	Jim Rice (3rd)
MVP (NL)	Bob Gibson (1st)	Pete Rose (6th)	Dave Parker (1st)
Cy Young (AL)	Denny McLain (4th)	Jim Palmer (13th)	Ron Guidry (1st)
Cy Young (NL)	Bob Gibson (1st)	Tom Seaver (1st)	Gaylord Perry (30th)
	1983	1988	1993
MVP (AL)	Cal Ripken Jr. (11th)	Jose Canseco (3rd)	Frank Thomas (3rd)
MVP (NL)	Dale Murphy (3rd)	Kirk Gibson (17th)	Barry Bonds (1st)
Cy Young (AL)	LaMarr Hoyt (21st)	Frank Viola (24th)	Jack McDowell (17th)
Cy Young (NL)	John Denny (14th)	Orel Hershiser (7th)	Greg Maddux (3rd)
	1998	2003	2008
MVP (AL)	Juan Gonzalez (18th)	Alex Rodriguez (7th)	Dustin Pedroia (23rd)
MVP (NL)	Sammy Sosa (7th)	Barry Bonds (1st)	Albert Pujols (1st)
Cy Young (AL)	Roger Clemens (3rd)	Roy Halladay (15th)	Cliff Lee (8th)
Cy Young (NL)	Tom Glavine (10th)	Eric Gagne (8th)	Tim Lincecum (1st)

Table 2
Player rankings. Top 10 pitchers (P) and batters (B) according to geodesic node betweenness (Btw), nestedness (N), and random walker ranking (R). Pitchers are divided into relief pitchers (RP) and starting pitchers (SP). In accordance with HOF eligibility, this table only includes players who played at least 10 seasons between 1954 and 2008. The random walker ranking values for batters, which we obtained from equation Eq. (2), are (in units of 10^{-5}): Bonds ≈ 9.22, Helton ≈ 7.94, Mantle ≈ 7.79, Ramirez ≈ 7.52, Thomas ≈ 7.15, Mays ≈ 6.90, McGwire ≈ 6.88, Rodriguez ≈ 6.86, Walker ≈ 6.85, and Guerrero ≈ 6.82. Note that if we consider all players with careers of at least 10 seasons, no matter how many of those seasons occurred between 1954 and 2008, the only change is that Ted Williams becomes the highest-ranking batter. If we consider all players with at least 8 seasons, the only additional change is that Albert Pujols is ranked just behind Barry Bonds.

Btw(P)	N(P)	R(RP)	R(SP)	Btw(B)	N(B)	R(B)
Nolan Ryan	Jamie Moyer	Mariano Rivera	Pedro Martinez	Julio Franco	Rickey Henderson	Barry Bonds
Jim Kaat	Roger Clemens	Billy Wagner	Roger Clemens	Rickey Henderson	Barry Bonds	Todd Helton
Tommy John	Greg Maddux	Troy Percival	Roy Halladay	Carl Yastrzemski	Steve Finley	Mickey Mantle
Dennis Eckersley	Mike Morgan	Trevor Hoffman	Curt Schilling	Hank Aaron	Craig Biggio	Manny Ramirez
Jamie Moyer	Randy Johnson	Tom Henke	Sandy Koufax	Pete Rose	Gary Sheffield	Frank Thomas
Greg Maddux	David Wells	B. J. Ryan	Randy Johnson	Tony Perez	Ken Griffey Jr.	Willie Mays
Charlie Hough	Kenny Rogers	Armando Benitez	John Smoltz	Joe Morgan	Luis Gonzalez	Mark McGwire
Don Sutton	Terry Mulholland	John Wetteland	Mike Mussina	Dave Winfield	Julio Franco	Alex Rodriguez
Phil Niekro	Jose Mesa	Keith Foulke	J. R. Richard	Ken Griffey Jr.	Jeff Kent	Larry Walker
Roger Clemens	Tom Glavine	Robb Nen	Greg Maddux	Al Kaline	Omar Vizquel	Vladimir Guerrero

4. Linking structure to performance

As we suggested previously, the network architecture should have important effects on the performance of players, as the quality of head-to-head competition is affected. In particular, central players in the network might have a systematic advantage in the rankings relative to those who are not as well-connected. Such structurally-important players (see Table 2 for examples), who have high values for both betweenness centrality and nestedness, have had long—and usually extremely successful—careers, so it is of significant interest (yet difficult) to gauge the coupled effects on their rank ordering from statistical success versus structural role in the network. With this in mind, we found almost no correlation ($\rho \approx 0.001$) between a player's position – i.e., individual nestedness and betweenness – and his success measured by the fraction of votes received. Accordingly, network position by itself (i.e., without also taking performance into account) does not seem to give any systematic advantage in a player's ranking.

To investigate this further, we examine the correlation between the sensitivity of rankings and changes in outcomes in individual pitcher–batter pairs. We do this by calculating the Moore–Penrose pseudo-inverse $\mathbf{L}^+$ of the graph Laplacian. Consider changing the outcome of the single edge that corresponds to the aggregate matchup between players i and j. If we increase the former's aggregate matchup by a unit amount at the expense of the latter, then the total change in votes $\mathbf{V}$ is proportional to the difference between the ith and jth columns of $\mathbf{L}^+$. This difference yields a node-centric measure of the sensitivities of rankings to individual performances. The constraint $\sum_i L^+_{ij} = 0$ implies that L^+_{ii} (the diagonal element of

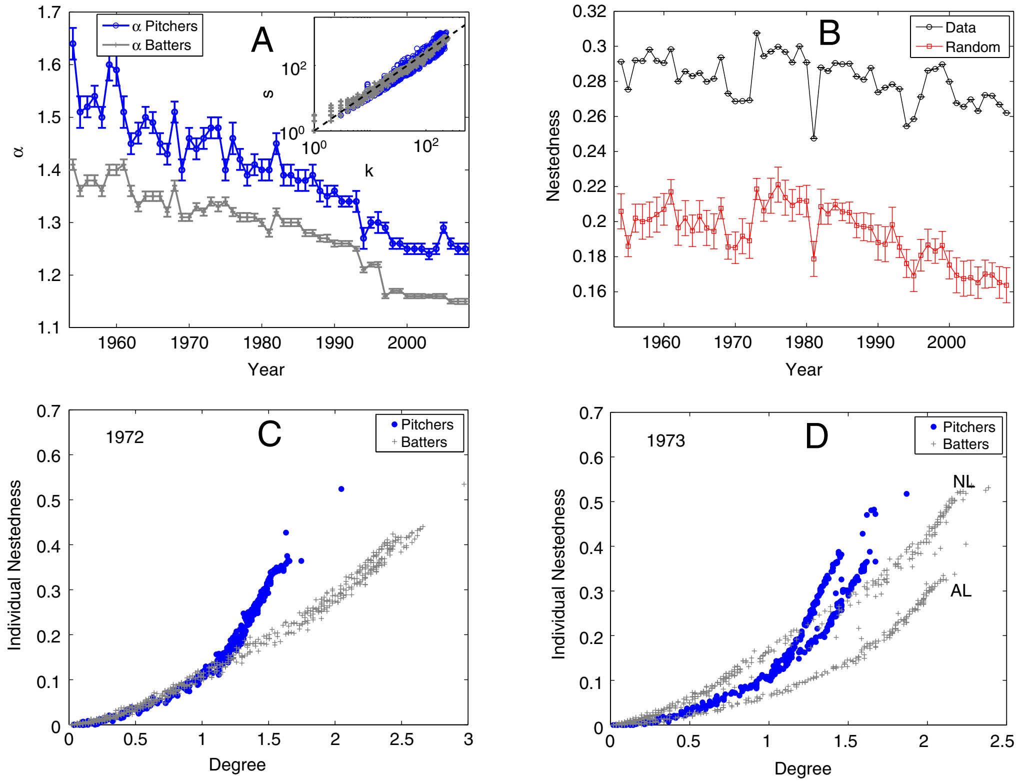

Fig. 3. (Color online) Time evolution and summary statistics of the baseball networks. Panel A shows the relation between player degree k and player strength s from 1954 to 2008. The vertical axis gives the value of the exponent α in the power-law relationship $s \sim k^{\alpha}$ (see the discussion in the main text), where we observe that α tends to decrease as a function of time. Shuffling the strengths in the network while keeping the player degrees fixed yields a power-law relationship with $\alpha \approx 1$ for all years. Blue circles denote pitchers and gray crosses denote batters. Each error bar corresponds to one standard deviation. The inset shows on a log–log scale the relationship between degree k and strength s for the 2008 season. Panel B shows the time evolution of the network's nestedness (which we defined using the NODF metric [26]). Black circles and red squares represent, respectively, the values for the original data and those for the standard null model II [21]. Each error bar again corresponds to one standard deviation. Panels C and D show, respectively, the relationship between node degree and individual nestedness for the 1972 and 1973 networks. For comparison purposes, the degrees of pitchers and batters are respectively scaled by a multiplicative factor of P/l and B/l, where P is the number of pitchers, B is the number of batters, and l is the number of undirected edges in the network. In 1973, the American League introduced the designated hitter rule, which caused a significant change in the structure of that and all subsequent single-season networks. Between 1954 and 1972, pitchers and batters each collapse onto a single curve. From 1973 to 2008, however, pitchers and batters each yield two distinct curves, revealing a division between the American league (bottom curve) and National League (top curve).

the pseudo-inverse of the graph Laplacian), which describes the direct control that player i has on his own ranking, is equal and opposite to the total change his performance directly imposes on the rest of the network. Additionally, as illustrated in Fig. 4A, the quantity L^+_{ii} is closely related to the total root-mean-squared (RMS) change in votes across the network due to the performance of player i. In particular, consider a player i on the right part of this figure. Such a player has low strength s_i and few appearances, so the largest value by far of the ith column of L^+ is $L^+_{ii} \approx 1/s_i$ (see the discussion below for more information on the relation to s_i). For instance, at the extreme right edge of the diagonal of L^+ for 2008, one finds Philadelphia Phillies middle reliever Geoff Geary (who we label as node g), who made an out in his only plate appearance and gives a diagonal entry of $L^+_{gg} \approx 1.0656$. Adding to Geary's charge in the network raises not only his value but also that of any pitcher that he faced. However, the value added to each of those pitchers is only about 0.0662. Hence, when we take the square root of the sum of squares, we obtain a value that is essentially indistinguishable from 1.0656. The asymptote on the right of Fig. 4A is thus the line $y = x$.

Note that the element L^+_{ii} is related to the mean of the commute distances between nodes i and j (averaging over all j) [35]. Given the constraint discussed above, the sum over j of the commute distances, given by $t_{ij} = L^+_{ii} + L^+_{jj} - 2L^+_{ij}$, yields a linear function of L^+_{ii}. Consequently, L^+_{ii} provides a node-centric measure of the average distance from node i to the rest of the network. The notion of average commute distance is reminiscent of the measures known as information centrality [36] and random walk centrality [37] (though the results of applying the different measures can still be quite different). The negative relationship between L^+_{ii} and both betweenness centrality and nestedness (see Fig. 5) thus yields a corresponding negative relationship between the mean commute distance and the betweenness and nestedness of a player. A player who is highly embedded in the network (i.e., one with high individual nestedness) has a small mean commute distance to the rest of the network, and the ranking of that player is not very sensitive to the outcome of a single matchup. In contrast, a player who is in the periphery of the network (i.e., one with low individual nestedness) typically has a very large mean commute distance to other portions of the graph, and his place in the rank ordering is consequently much more sensitive to the results of his

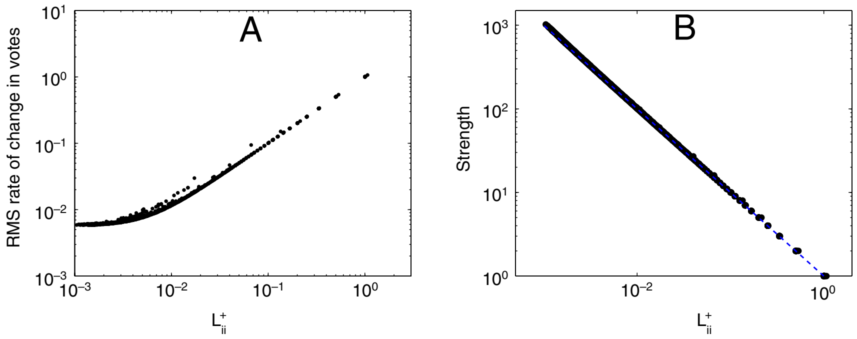

Fig. 4. (Color online) Network quantities versus graph Laplacian. We plot the diagonal elements L^+_{ii} of the Moore–Penrose pseudo-inverse of the graph Laplacian for the 2008 baseball network versus (A) the root-mean-squared (RMS) change of votes across the network due to the RUE 'charge' at each node and (B) node strength. In each case, we use logarithmic coordinates on both axes. (In this example, we keep track of all players regardless of number of appearances.) The plateau that we observe in the left of panel A is unsurprising, as it corresponds to a bevy of high-strength players with a large number of interconnections—i.e., what seems to be a core structure in the nested network—so the total RMS effects appear to be independent of the increase in charge on any particular core player. We discuss the asymptotic behavior on the right of panel A in the main text and note the $L^+_{ii} \approx s_i^{-1}$ relationship in panel B. We observe similar properties in the other seasons.

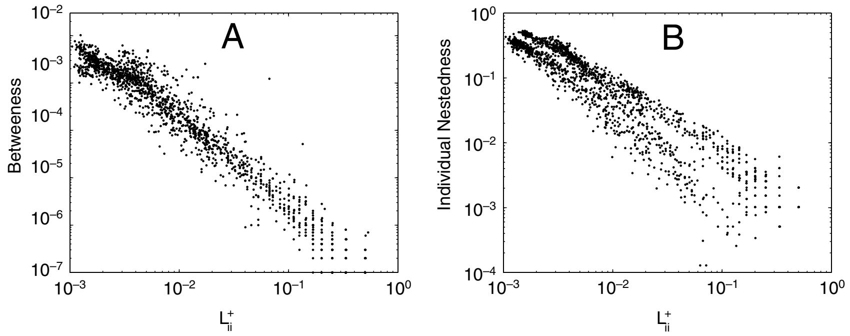

Fig. 5. (Color online) Betweenness and Nestedness versus Graph Laplacian. We plot the diagonal elements of the Moore–Penrose pseudo-inverse of the graph Laplacian for the 2008 baseball network versus (A) node betweenness and (B) individual nestedness. (In this example, we keep track of all players regardless of number of appearances.) The two sets of points in Panel B reveal two subsets of the data (namely, batters and pitchers). We observe similar properties in the other seasons.

individual matchups [38]. This suggests that players in the AL tend on average to be very slightly more prone to changes in their individual matchup outcomes than players in the NL (see Fig. 3D).

Remarkably, we can make these general notions much more precise, as $L^+_{ii} \approx s_i^{-1}$, where we recall that s_i is the strength of node i (see Fig. 4B). Some similarities between these quantities is expected. (Consider, e.g., the role of relaxation times in a similar relationship—which can be quantified via an eigenvalue analysis—with random walk centrality in Ref. [37].) The accuracy of this simple relationship belies a stunning organizational principle of the baseball matchup networks: The global quantity of average commute distance of a node is well approximated by its strength, a simple local quantity. That is, in the appropriate perturbation analysis to approximate the pseudo-inverse of the Laplacian, the higher-order terms essentially cancel out, contributing little beyond the (zeroth-order) local contribution. We also found a rougher relationship for nestedness and betweenness (see Fig. 6).

These results have two interesting implications. First, they reveal that the success of well-connected players depends fundamentally on a strong aggregate performance rather than simply on their position in the network. Second, they imply that neophyte players would need to face well-connected players if they want to establish a stronger connection to the network and a ranking that is less vulnerable to individual matchups. Similarly, recent research on mutualistic networks in ecology has found that neophyte species experience lower competition pressures by linking to well-connected species [9]. Our findings on baseball player rankings suggest the possibility that one might find similar competition patterns in mutually-antagonistic interactions in ecological and social networks.

5. Conclusions

Drawing on ideas from network science and ecology, we have analyzed the structure and time evolution of mutually-antagonistic interaction networks in baseball. We considered a simple ranking system based on biased random walks on these graphs and used it to compare player performance in individual seasons and across entire careers. We emphasize that our ranking methodology is overly simplistic, as there are several considerations that one might use to improve it (see, e.g., Appendix B) while maintaining a network framework that accounts for the opponents that each player has faced. We also examined how the player rankings and their sensitivities depend on node-centric network characteristics.

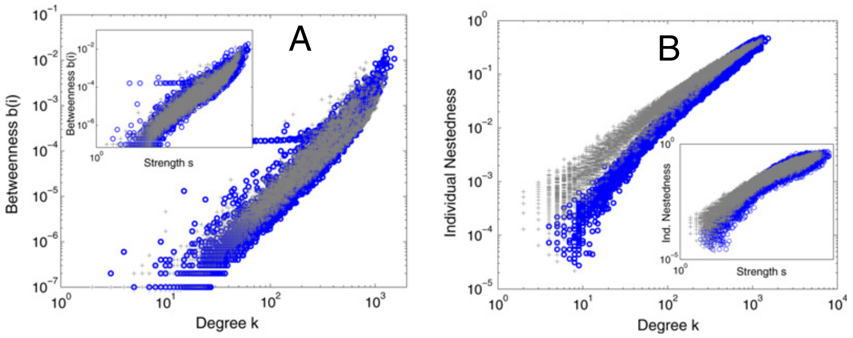

Fig. 6. (Color online) Degree, Strength, Betweenness, and Nestedness. We show a log–log plot of (A) player degree k versus node betweenness centrality and (B) degree versus individual nestedness in the career network. The insets show the analogous relationships obtained by replacing degree with strength s. Pitchers are shown by blue dots and batters are shown by gray crosses. Pitchers with betweenness $b \approx 2 \times 10^{-4}$ and low degree tend to be position players who made a few pitching appearances (e.g., Keith Osik), pitchers with short careers (e.g., Wascar Serrano), or recent pitchers with few Major League appearances (e.g., John Van Benschoten, who has split time between the Major Leagues and the Minor Leagues since 2004).

We expect that similar considerations might be useful for developing a better understanding of the interplay between structure and function in a broad class of competitive networks, such as those formed by antigen–antibody interactions, species competition for resources, and company competition for consumers. We are optimistic that this might lead to interesting ecological insights, perhaps by compensating for the difficulty in collecting data on the regulatory dynamics of mutually-antagonistic networks in ecology—such as the ones formed by parasites and free-living species [11]—or by helping to assess the potential performance of invasive species from different environments [39].

Acknowledgements

We thank Eduardo Altmann, Jordi Bascompte, Mariano Beguerisse Díaz, Damian Burch, Justin Howell, Peter Jensen, Tom Maccarone, Ravi Montenegro, Cy Morong, Felix Reed-Tsochas, David Smith, Daniel Stouffer, Tom Tango, and Andrew Thomas for their useful comments. SS did some of his work while he was a Postdoctoral Fellow at the Oxford University Corporate Reputation Centre and the CABDyN Complexity Centre. PJM, SP, and TM were funded by the NSF (DMS-0645369) and by start-up funds to PJM provided by the Institute for Advanced Materials, Nanoscience and Technology and the Department of Mathematics at the University of North Carolina at Chapel Hill. MAP acknowledges a research award (#220020177) from the James S. McDonnell Foundation. We obtained the baseball network data from retrosheet.org. The data was obtained free of charge from and is copyrighted by Retrosheet (20 Sunset Rd., Newark, DE 19711, USA).

Appendix A. Quantities for bipartite networks

In this appendix, we review some important quantities for bipartite networks and discuss their values for the baseball matchup networks.

A clustering coefficient for unweighted bipartite networks can be defined by [40]

$$C_{4,mn}(i) = \frac{q_{imn}}{(k_m - \eta_{imn}) + (k_n - \eta_{imn}) + q_{imn}}, \tag{A.1}$$

where q_{imn} is the number of complete squares involving nodes i, m, and n; the quantity $\eta_{imn} = 1 + q_{imn}$ enforces the requirement in bipartite graphs that there are no links between nodes of the same population; and we recall that k_i is the degree of node i. Hence, the numerator in (A.1) gives the actual number of squares and the denominator gives the maximum number of possible squares. For the single-season unweighted matchup networks, we calculate the ratio $r_c = \langle C_4 \rangle / \langle C_{4r} \rangle$ between the mean clustering coefficient $\langle C_4 \rangle$ summed over all nodes i and the mean clustering coefficient $\langle C_{4r} \rangle$ generated by a randomization of the network that preserves the original degree distribution [41]. We find that the unweighted matchup networks have average clustering coefficients that are just above that of random networks. Interestingly, the ratio r_c decreases gradually (and almost monotonically from one season to the next) from $r_c \approx 2.5$ in 1954 to $r_c \approx 1.3$ in 2008.

The geodesic betweenness centrality of nodes over the unweighted network $\hat{\mathbf{A}}$ is defined by [1,42]

$$b(i) = \sum_{j,k} \frac{\Delta_{j,k}(i)}{d_{j,k}}, \tag{A.2}$$

where $\Delta_{j,k}(i)$ is the number of shortest paths between players j and k that pass through player i and $d_{j,k}$ is the total number of shortest paths between players j and k. For the single-season baseball networks, we calculate the ratio $r_b = \langle b \rangle / \langle b_r \rangle$ between the mean path length $\langle b \rangle$ summed over all nodes i and the mean path length $\langle b_r \rangle$ generated by a randomization of the network that preserves the degree distribution [41]. As with clustering coefficients, we find that the mean path lengths of baseball networks are only slightly larger than those of random networks, finding in particular that $r_b \in (1, 3)$.

Nestedness is an important concept that has been applied to ecological communities, in which species present in sites with low biodiversity are also present in sites with high biodiversity [43]. Although the general notion of nestedness can vary, the concept has nonetheless been employed quite successfully in the analysis of ecological networks [21]. In a nested network, interactions between two classes of nodes (e.g., plants and animals) are arranged so that low-degree nodes interact with proper subsets of the neighbors of high-degree nodes. A nested network contains not only a core of high-degree nodes that interact with each other but also an important set of asymmetric links (i.e., connections between high-degree and low-degree nodes). The importance of nestedness measures is twofold: (1) they give a sense of network organization; and (2) they have significant implications for the stability and robustness of ecological networks [9,21].

To avoid biases in nestedness based on network size (i.e., the number of nodes), degree distribution, and other structural properties, we employ the nestedness calculations introduced recently in Ref. [21]. The aggregate nestedness (of an unweighted matchup network) is given by [26]

$$NODF = \frac{\sum_{i,j} N_{i,j} + \sum_{l,m} N_{l,m}}{([P(P-1)/2] + [B(B-1)/2])}. \tag{A.3}$$

For every pair of pitchers (i and j), the quantity $N_{i,j}$ is equal to 0 if $k_i \leq k_j$ and is equal to the fraction of common opponents if $k_i > k_j$. We also define a similar quantity for every pair of batters (l and m). The nestedness metric takes values in the interval $[0, 1]$, where 1 designates a perfectly-nested network and 0 indicates a network with no nestedness.

The NODF version of nestedness also allows one to calculate the individual nestedness of each pitcher (column) or batter (row) using the equation

$$z(i) = \sum_j N_{i,j}/(T-1), \tag{A.4}$$

where $T = P$ (total number of columns) for pitchers, $T = B$ (total number of rows) for batters, and $N_{i,j}$ is calculated as above. In this way, the individual nestedness metric takes values in the interval $[0, 1]$, where 1 designates a perfectly-nested individual and 0 indicates an individual with no nestedness.

The null model used to provide a comparison for the nestedness calculation is [21]

$$q_{i,j} = \frac{k_i}{2B} + \frac{k_j}{2P}, \tag{A.5}$$

where $q_{i,j}$ is called the occupation probability of an interaction between node i and node j and we recall that B and P are, respectively, the total number of nodes j (batters) and nodes i (pitchers) in the network. In a bipartite network, j and i represent two different types of nodes, so $q_{i,j}$ is the mean of the occupation probabilities of the row and column. For our study, we also calculate the standard error—given by $Z = (NODF - \langle NODF \rangle)/\sigma$, where $NODF$ corresponds to the nestedness values of the empirical networks and $\langle NODF \rangle$ and σ are, respectively, the average and standard deviations of nestedness values of random replicates generated by the null model. For the unweighted matchup networks, we find that $Z > 3$ in all seasons (see Fig. 3B).

Appendix B. Definition of Runs to End of Inning (RUE)

To quantify the outcome of each plate appearance, we used the sabermetric quantity *runs to end of inning* (RUE) [15], which assigns a value to each of the possible outcomes in a plate appearance based on the expected number of runs a team would obtain before the end of that inning following that event, independent of game context. (RUE can also be adjusted by subtracting the initial run state [34].) Higher numbers indicate larger degrees of success for the batter. The possible events (and their associated numerical RUE values) are as follows: generic out (0.240), strikeout (0.207), walk (0.845), hit by pitch (0.969), interference (1.132), fielder's choice (0.240), single (1.025), double (1.311), triple (1.616), and home run (1.942).

Note that we are ignoring events such as passed balls and stolen bases that can occur in addition to the above outcomes in a given plate appearance. This might lead to some undervaluing in the ranking for a small number of position players (such as Tim Raines) that rely on stolen bases. We also considered the metric known as weighted on base average (wOBA) [44] and note that any metric that assigns a value to a specific plate appearance can be used in place of RUE without changing the rest of our ranking algorithm. This includes, in particular, popular sabermetric quantities such as win shares and value over replacement player (VORP) [13,15]. One can also incorporate ideas such as ballpark effects into the metric employed at this stage of the algorithm without changing any other part of the method. Although it would make the methodology more complicated (in contrast to our goals), it is also possible to generalize our algorithm to include more subtle effects such as estimates for when player performance peaks and how it declines over a long career. Some of the active players in the data set have not yet entered a declining phase in their careers and might have higher rankings now than they will when their careers are over. We expect that the relatively high rankings of modern players versus ones who retired long ago might also result in part from the increased performance discrepancy between the top players and average players in the present era versus what used to be the case and in part from performing well against the larger number of relatively poor players occupying rosters because of expansion [45]. Finally, we note that batter–pitcher matchups are not fully random but contain significant correlations (e.g., in a given baseball game, the entire lineup of one team has plate appearances against the other team's starting pitcher) that can be incorporated to generalize the random walker process itself [34].

To include the outcomes of players who did not have many plate appearances without skewing their rankings via small samples, we separately accumulated the results for all pitchers and batters with fewer than some threshold number of plate appearances K into a single "replacement pitcher" and "replacement batter" to represent these less prominent players. In the results presented in this paper, we used the threshold $K = 500$ for both the single-season and career networks. Note that similar thresholds exist when determining single-season leadership in quantities such as batting average (which requires 3.1 plate appearances per team game, yielding 502 appearances in a 162-game season) and earned run average (which requires 1 inning per team game).

References

[1] M.E.J. Newman, SIAM Rev. 45 (2003) 167.
[2] G. Caldarelli, Scale-Free Networks: Complex Webs in Nature and Technology, Oxford Univ. Press, 2007.
[3] M.A. Porter, J.-P. Onnela, P.J. Mucha, Notices Amer. Math. Soc. 56 (2009) 1082.
[4] R.M. May, Nature 238 (1972) 413.
[5] M. Pascual, J. Dunne (Eds.), Ecological Networks: Linking Structure to Dynamics in Food Webs, Oxford Univ. Press, 2005.
[6] U. Bastolla, M. Lassig, S.C. Manrubia, A. Valleriani, J. Theor. Biol. 235 (2005) 521.
[7] S. Saavedra, F. Reed-Tsochas, B. Uzzi, Proc. Natl. Acad. Sci. 105 (2008) 16466.
[8] S. Saavedra, F. Reed-Tsochas, B. Uzzi, Nature 457 (2009) 463.
[9] U. Bastolla, M.A. Fortuna, A. Pascual-García, A. Ferrera, B. Luque, J. Bascompte, Nature 458 (2009) 1018.
[10] T.L. Czárán, R.F. Hoekstra, L. Pagie, Proc. Natl. Acad. Sci. 99 (2002) 786.
[11] K.D. Lafferty, A.P. Dobson, A.M. Kuris, Proc. Natl. Acad. Sci. 103 (2006) 11211.
[12] Retrosheet has pre-1954 data, but because that data is not currently complete, we do not extract pitcher–batter matchups from earlier years.
[13] B. James, What Happened to the Baseball Hall of Fame, Fireside, 1995.
[14] J. Thorn, P. Palmer, The Hidden Game of Baseball: A Revolutionary Approach to Baseball and its Statistics, Garden City, 1984.
[15] B. James, Win Shares, STATS 2002.
[16] S.M. Berry, C.S. Reese, P.D. Larkey, J. Amer. Stat. Assoc. 94 (1999) 661.
[17] Of course, the decisions of the HOF voters are influenced not only by perceptions of player performance in batter–pitcher matchups but also by perceptions of other factors such as defense, leadership/attitude/clubhouse presence, clutch performance, use of performance-enhancing drugs, etc. The current list of HOF members, players who will shortly become eligible for membership, and past voting results are available on the National Baseball Hall of Fame and Museum's website (www.baseballhalloffame.org).
[18] T. Callaghan, M.A. Porter, P.J. Mucha, Notices Amer. Math. Soc. 51 (2003) 887.
[19] T. Callaghan, M.A. Porter, P.J. Mucha, Amer. Math. Monthly 114 (9) (2007) 761.
[20] J.L. Gross, J. Yellen, Handbook of Graph Theory, CRC Press, 2004.
[21] J. Bascompte, P. Jordano, C.J. Melián, J.M. Olesen, Proc. Natl. Acad. Sci. 100 (2003) 9383.
[22] P. Dickson, The Dickson Baseball Dictionary, third ed., W. W. Norton & Comp, 2009.
[23] A.M. Petersen, W.-S. Jung, H.E. Stanley, Europhys. Lett. 83 (2008) 50010.
[24] J. Bascompte, P. Jordano, J.M. Olesen, Science 312 (2006) 431.
[25] E. Ben-Naim, F. Vazquez, S. Redner, J. Quant. Anal. in Sports 2 (4) (2006) 1.
[26] M. Almeida-Neto, P. Guimarães, P.R. Guimarães Jr., R.D. Loyola, W. Ulrich, Oikos 117 (2008) 1227.
[27] N. Silver, Introducing PECOTA, Brassey's Publishers, 2005.
[28] The Hardball Times Writers, The Hardball Times Baseball Annual 2009, The Hardball Times, 2008.
[29] J.P. Keener, SIAM Rev. 35 (1993) 80.
[30] A.N. Langville, C.D. Meyer, SIAM Rev. 47 (2005) 135.
[31] D. Marincic, Todd Helton, the Hall of Fame, and a rain-delayed White Sox game, in: NL Central Stage blog, http://nlcentralstage.blogspot.com/2009/06/todd-helton-hall-of-fame-and-rain.html (13 June 2009).
[32] As we discuss in Section B, we did not incorporate ballpark effects into our ranking methodology
[33] ESPN, http://espn.go.com/mlb/playerratings/_/year/2008.
[34] T. Tango, Six degrees of WOWY, in: THE BOOK–Playing the Percentages in Baseball blog, http://www.insidethebook.com/ee/index.php/site/article/six_degrees_of_wowy/ (5 August 2009).
[35] D. Aldous, J.A. Fill, Reversible Markov Chains and Random Walks on Graphs (in preparation) Available at: www.stat.berkeley.edu/~aldous/RWG/book.html.
[36] K. Stephenson, M. Zelen, Social Networks 11 (1989) 1.
[37] J.D. Noh, H. Rieger, Phys. Rev. Lett. 92 (2004) 118701.
[38] This feature is a sensible one, as players with high nestedness values have typically accumulated many more plate appearances—whether as a regular starter in a single-season network or over a long career in the multi-season network—relative to players who are further in the periphery of the network.
[39] F. Bulleri, J.F. Bruno, L. Benedetti-Cecchi, PLoS Biol. 6 (2008) e162.
[40] P. Zhang, J. Wang, X. Li, M. Li, Z. Di, Y. Fan, Physica A 387 (2008) 6869.
[41] S. Maslov, K. Sneppen, Science 296 (2002) 910.
[42] L.C. Freeman, Sociometry 40 (1977) 35.
[43] W. Atmar, B.D. Paterson, Oecologia 96 (1993) 373.
[44] T. Tango, M. Lichtman, A. Dolphin, The Book: Playing the Percentages in Baseball, Protomac Books, 2007.
[45] S. Forman, Baseball-reference.com (http://www.baseball-reference.com/).