Fixing Flaws and Stopping Draws*

A new chess tie-break system based on directed network analysis

David Smerdon

June 30, 2012

1 Introduction

The 2011 European Chess Championships saw over 400 of Europe's top chess players compete for one of 23 qualifying spots for the World Chess Cup, and with it a shot at the world title and the €1.5 million prize fund. With only eleven rounds played in the Swiss-system tournament (where all players play all eleven games), the chances of potential qualifiers being tied on the same score were high, and so it transpired: Four players finished on 8.5 points (out of eleven)[1], eleven players finished with 8.0, and a further 29 finished on 7.5. The eventual decision of how to decide the final eight available qualifying positions from these 29 players was highly controversial, leading to official protests, heated debate in all levels of the chess community from official international forums to amateur clubs, and eventually an admission from FIDE[2] that the current tie-breaking regulations were woefully inadequate.

At the same time, in South Africa, the 2011 Commonwealth Chess Championships also ended in a tie, on this occasion for first place. Despite the employment of the official FIDE tie-break procedures in this eleven-round event with over 700 players, the gold medal was ultimately decided by the result on board 44 of by a player who finished 144th. It has become increasingly clear that current tie-break methodologies in large tournaments are proving unsatisfactory in their stated objective of determining the strongest performing player.

Meanwhile, an escalating issue in international chess over the past few years is the high number of draws (and particularly 'soft' draws) between top players. With typically several games played each day, each lasting up to five hours, open tournaments are acutely mentally draining. As a result, many grandmasters employ a strategy of agreeing to quick draws in games against other grandmasters to conserve energy. This helps to ensure they win their games against weaker players, frequently netting a score sufficient for a large expected payout from the prize fund. However, the consequences for spectators, sponsors and the sport

*Social Network Analysis Research Proposal, Tinbergen Institute

[1]In international chess, a win is recorded as one point, a loss as zero, and a draw earns a player 0.5 points.

[2]The Fédération Internationale des Échecs, or World Chess Federation

in general are rather more damaging. In an age when grandmasters are concerned about protecting their livelihood but the general public craves increasing levels of excitement and drama for their entertainment purposes, professional chess is at risk of becoming obsolete. Many radical measures have been suggested to combat the draw problem in chess, but none have found favour with the overarching chess community, traditionalists and liberals alike.

This proposal seeks to address both the inadequacies of current tie-break systems and the issue of ever-increasing draws in chess by introducing a new tie-break system for large tournaments. The system uses a measure based on both direct and indirect wins, and is a generalisation of standard centrality values stemming from the directed network of wins throughout an event.

2 Background

2.1 Chess tie-break systems

The above examples are not unusual in the chess world or indeed in many other sporting and gaming contests. This is particularly the case in large open tournaments in which players typically play against different opponents and in which it may be undesirable or impractical for ties to be broken by straight head-to-head playoffs. In the majority of chess events held around the world, the so-called Swiss pairing system matches players of a similar score against each other over a fixed number of rounds (usually seven, nine or eleven), with the top finishers usually facing opponents of increasing difficulty over the course of the competition. With participation numbers ranging from around twenty to several hundred, ties are an inevitability.

Acknowledging the lack of one unanimously accepted or 'best' system, FIDE permits the organisers of chess events to pick from a list of officially sanctioned tie-break systems for Swiss system tournaments. A brief summary follows.

The following four systems are relatively rare and generally considered impractical or even methodologically flawed for large open tournaments:

- ***Direct encounter:*** The result of the game between two tied players decides. This measure fails to split the tie if the players were not matched in the tournament, or if they were and the result was a draw, or if there are more than two tied players and a directed cycle of wins (or purely draws) among the tied players exists.
- ***Playoff:*** These are typically played at quicker, 'blitz' time controls. With open tournaments usually organised on a tight schedule over a few days, including a closing ceremony, this is often logistically undesirable due to time constraints. This pressure is amplified in case the playoff ends in a draw, or again in case of multiple tied players.

- ***Games played with Black:***[3] Given that in chess White has the first move and thus (at least ostensibly) a slight advantage, this measure favours the player deemed to have been disadvantaged in the random colour assignment. However, given that a strong player is highly likely to beat the weaker opponents in the earlier rounds with either colour, this system is usually considered akin to a coin flip in deciding major prizes.

- ***Tournament Performance Rating:*** This is a measure of performance using the existing FIDE ratings of a player's opponents. It thus does not account for opponents' performances in the actual event, and is prone to underweighing opponents who are 'underrated', and vice-versa.

The following four systems are by far the most widely used by organisers:

- ***Won games:*** Only the number of wins is counted. This measure is designed to encourage 'fighting chess', represented by a higher proportion of decisive results, and to discourage the 'Grandmaster draw' mentality (see above).

  The main criticism with this system is that it favours players who play badly in early rounds and are then paired against much weaker opponents against whom they subsequently win (the so-called 'Yoyo Effect'), compared to players who made two quality draws against strong opposition[4].

- ***Sum of progressive score:*** The scores of the player after each round are cumulatively added, thus weighting earlier rounds higher.

  This aims to circumvent the 'Yoyo Effect'[5]. However, it disproportionately favours higher seeds (those with higher preexisting FIDE ratings), as they face far easier opposition in early rounds than the rest of the field.

- ***Buchholz:*** The sum of the score of each of the opponents of a player is used to represent the strength of a player's opposition. Derivations of the system may exclude the score of the lowest-rated or highest-rated opponent, or both.

  This is currently the most popular system, and was used in both the 2011 Commonwealth Championships and European Championships (in combination with Tournament Performance Rating in the latter case). However, when contenders at the top meet in later rounds, it actually encourages quick draws in order to maximise their expected payoff. In addition to conserving energy, quick draws among strong players reduce the risk of a poorer tie-break score resulting from losing and having to 'yoyo' upwards.

- ***Sonneborn-Berger:*** Similar to Buchholz, but using the sum of the scores of the opponents a player

[3] As tournaments are usually played over an odd number of rounds, players will typically have played one more game with one colour than with another.

[4] In the Swiss pairing system, players are paired each round against other players of the same or similar score. Thus, a player who scores poorly in the early games and catches up to the top finishers at the end of the event will, on average, have played a weaker field. In chess parlance, a strong player who unexpectedly loses a game in the beginning of an event but eventually wins a major prize due to having faced easier opposition is said to have employed the 'Swiss Gambit'.

[5] For example, two draws in succession yields a progressive tie-break score of $(0.5 + 1.0) = 1.5$, while a loss followed by a win gives $(0.0 + 1.0) = 1.0$.

has defeated and *half* the scores of the opponents a player has drawn with.

While this system may appear to encourage individual game victories, the measure's estimate of overall opponent strength places disproportionate emphasis on the performance of a player's weakest opponents, as they are the ones most likely to have been beaten. In that sense, a player again benefits from the 'Yoyo Effect'.

It is evident that no system is without its flaws. This explains FIDE's decision to leave the choice of system up to the organisers and arbiters of each event, without guidance or preference. Tournament organisers thus have a difficult choice to make, weighing up the typically conflicting criteria of *Fairness* (picking the player who played against the strongest field of opponents), *Combativeness* (picking the player who played the most 'fighting' chess) and *Practicality* (overcoming logistics; in this sense, any measure that can be computed without the need for additional games).

The Buchholz system, on the whole, performs better than its rivals in terms of *Fairness*, and is evidently practical. However, the example mentioned at the outset of the 2011 Commonwealth Championships demonstrates its limitations in a large participant environment. Furthermore, the exhaustive post-mortem of the 2011 European Championships also showed that employing a pure Buchholz system could give widely differing rankings on the basis of one or two 'irrelevant' games played between players not directly involved in the ties. And, as outlined above, Buchholz amplifies the perverse incentives for quick draws, increasing the (arguably unexciting) draw ratio and decreasing *Combativeness*. Out of the fifteen games played among the top fifteen place-getters in the European Championships, twelve (80%) were drawn. (Out of the roughly six million game results recorded from chess tournaments in the modern age, the draw ratio is approximately 38%.)

2.2 Draws in chess

While draws in chess occur a little over a third of the time on average, this statistic increases for players strong enough to make a living from the sport. Among strong club players the draw likelihood is only slightly higher, but approaches 50% for grandmasters and reaches above 60% for the world's top 100 players. In the recent 2012 World Championship match, ten of the twelve games between the two contenders were drawn, with only one of these taking longer than half the allotted time.

One might argue that this is a reflection of the quality of play, perhaps due to less blunders, so a dynamic comparison of draw likelihoods for top players over time is sensible. Sonas (2009[13]) found that the likelihood of draws in top-level chess has steadily increased since the start of the twentieth century, reaching historically its highest rate at present (see Figures 1 and 2). This is logical given the steadily increasing prizemoney in chess. For professionals playing the grueling international circuit, an energy-saving quick draw against another professional, in a game in which the expected result is in any case a draw, could well be an optimal strategy. This is a very topical and unresolved issue at present in chess administration, to the point that many radical solutions have been proposed. These include switching to a football-style scoring system of three points for a win, one for a draw and zero for a loss, and even modifying the laws of chess to exclude the possibility of a draw. However, chess is an ancient game steeped in tradition, and this research will not consider alternatives that alter the established scoring conventions or the structure of the game itself.

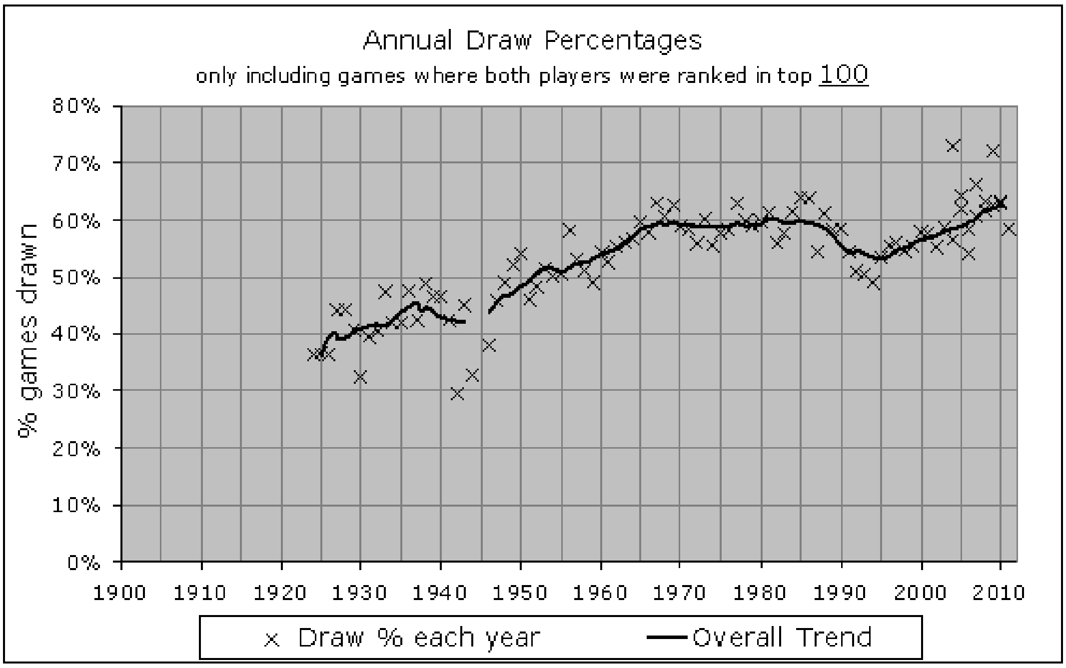

Figure 1: (Source: Sonas, 2011)

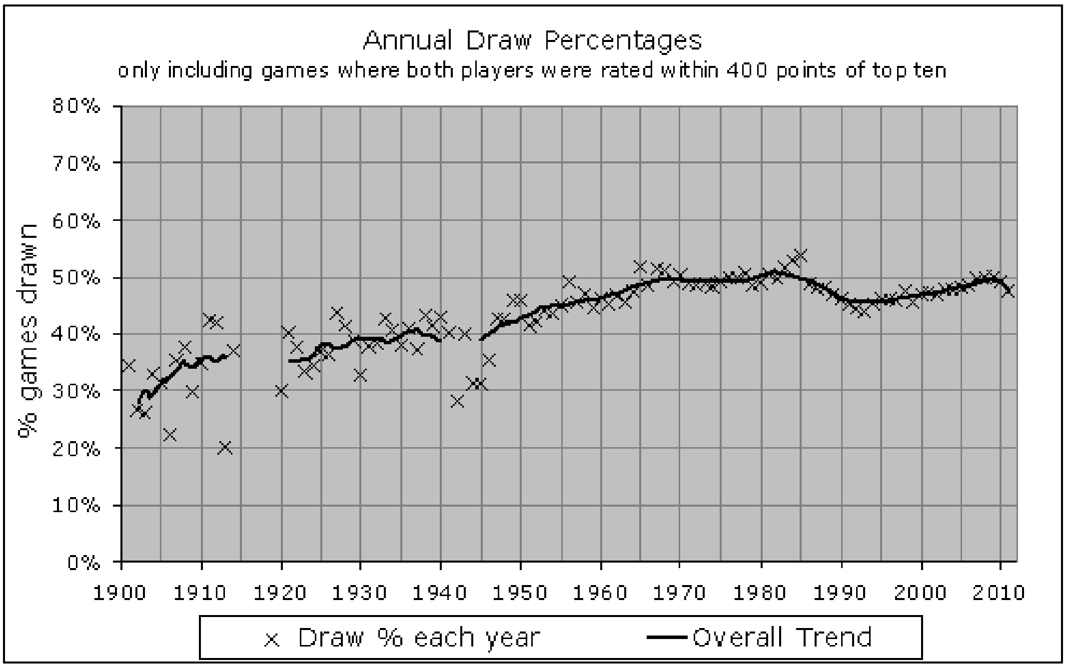

Figure 2: (Source: Sonas, 2011)

2.3 Proposed contribution

This proposal, instead, seeks to introduce a fair, combative and practical tie-break system to address the aforementioned issues by using network theory. To understand the motivation for this, observe that the issue of accounting for different opponent strength disappears in a round-robin (or 'all-play-all') system. In this instance, tied players have played against exactly the same opposition, and so the accepted tie-break system of *Direct Encounter* followed by *Won Games* is sensible and reasonably uncontroversial. However, for the vast majority of open tournaments, the $\frac{1}{2}n(n-1)$ games required for a round-robin with n participants is naturally impractical. It is thus impossible to compare the direct results by two tied players against all other participants.

However, it may be possible to compare the *indirect* results of tied players against all other participants. In fact, in the vase majority of chess tournaments each player is connected to each other player through their results, and this can be achieved with a very small number of 'links' (or rounds). This proposal seeks to investigate the merits of a tie-break system that considers the number of *weighted, indirect wins* as its comparison measure. It is proposed that such a system is both *Fair* and *Practical*, while encouraging *Combativeness* through its focus on victories.

3 Related Literature

As can be expected, the academic literature on the topic of chess tie-break systems is sparse. In his analysis of croquet tournaments, Appleton (1995[2]) notes that the issue of ties is the chief drawback of the Swiss system, and proposes yet another modification of the Buchholz system as a preferred tie-breaker for addressing *Fairness*. Butler (1997[4]) shows that rankings based on the Buchholz system become increasingly susceptible to large variations from small perturbations in 'insignificant' results by tied players' opponents as the number of participants grows. He suggests a combination of a player's final (or 'direct') score and the tie-break score in determining the final rankings to avoid ties, but we will not consider approaches that alter the fundamental chess scoring system.

There are a number of non-academic articles from the chess community analysing the issue of soft draws. Leong and Weiwen (2005[9]) suggest that a game ending in a draw should immediately be resolved by a so-called Armageddon blitz game[6]. Kasimdzhanov (2011[7]) suggests an expansion on this concept, but both approaches are unsuitable and impractical for the majority of non-elite tournaments. Perhaps the least intrusive approach, though somewhat subjective and prone to a Prisoners' Dilemma problem, comes from Nunn (2005[10]), who suggests that tournament organisers should collude to extend invitations to only those top players exhibiting "fighting spirit."

Closest to this proposal in the literature is a study on scoring systems in American college football by Park

[6]In this case, White has 6 minutes of thinking time and Black has 5 minutes, but a draw is recorded as a win for Black.

and Newman (2005[11]). This article seeks to create a new scoring system to rank teams despite the lack of a completely connected network in this football league. While this is a primary rather than tie-break scoring system and no draws are permitted in American football, the authors do employ similar methodology to that proposed below. Direct and indirect wins and losses are calculated, with the final ranking depending on the difference between the total weighted win- and loss-scores over the network.

The current research will adopt a similar approach, making use of standard network techniques for calculating centrality and betweenness measures by Katz (1953[8]), Bonacich(1991[3]) and Freeman et al. (1991[5]), which are contrasted and explained in greater depth in Jackson(2010[6]).

4 The Model

Consider a tournament with n players and r rounds, with $n > r$ [7]. Let G be the *directed graph* of wins for the tournament, with an out-arrow from i to j indicating that player i beat player j. Then the $n \times n$ adjacency matrix is A, with players $i = 1, ..., n$ corresponding to the rows, and correspondingly the columns $j = 1, ..., n$ represent when a player is in the role of the opponent. Let an entry A_{ij} be 1 if player i beat player j, and zero otherwise, with $A_{ii} = 0 \; \forall \; i$.

Then:

$$\sum_j A_{ij} = d_i$$

i.e the degree of node i gives the number of wins recorded by player i in the tournament. Let d being the vector containing the number of wins for each player.

We can think of these as *direct wins*, or alternatively, 'wins of length 1'. Now consider an *indirect win* of length 2 by i over k to be such that i beats j, j beats k for some player j (Figure 3).

The total number of indirect wins by player i of length 2 is:

$$\begin{aligned} d_{i,2} &= \sum_j A_{ij} (\sum_k A_{jk}) \\ &= \sum_j \sum_k A_{ij} A_{jk} \end{aligned}$$

To claim that player i would beat player k in a head-to-head game on this basis is somewhat tenuous, so intuitively it seems logical to weigh these wins less than those of length one. Define δ as the (to be determined) discount factor, and T_i as the total tie-break score. Then:

$$\begin{aligned} T_i &= d_i + \delta d_{i,2} + \delta^2 d_{i,3} + ... \\ &= \sum_j A_{ij} + \delta \sum_j \sum_k A_{ij} A_{jk} + \delta^2 \sum_j \sum_k \sum_l A_{ij} A_{jk} A_{kl} + ... \end{aligned}$$

[7]...and usually substantially so, by a factor of around ten in most chess tournaments.

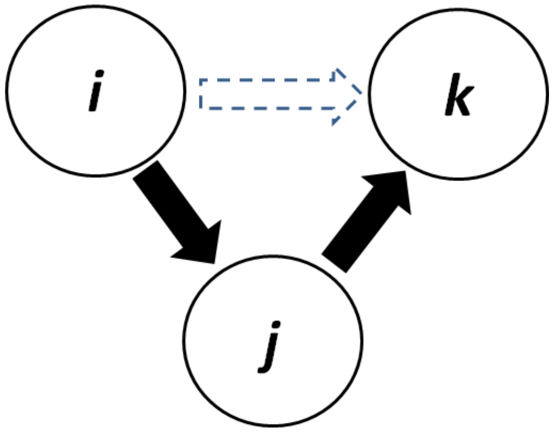

Figure 3: i directly beats j and indirectly beats k.

So:

$$T_i = d_i + \delta \sum_j A_{ij} T_j \tag{1}$$

That is, the tie-break score for player i is the number of direct wins plus the sum of the tie-break scores of player i's defeated opponents, discounted by one link.

From this equation, the two chief differences of this measure with the Buchholz system are clear:

Draws are not credited. That is, only wins from the player's direct score contribute to the tie-break. This follows the principle of *Combativeness*.

Wins against stronger opponents count higher. The summation of indirect wins can be seen to reflect a measure of the strength of the player's opponents, in line with Buchholz. However, the nature of the directed 'wins' network means that, on average, opponents who score more points in the tournament contribute more to the tie-break score (*Fairness*).

The vector of all tie-break scores, T, is then:

$$T = d + \delta A \cdot T$$

...which can be solved as:

$$(I_n - \delta A) \cdot T = d$$

$$T = (I_n - \delta A)^{-1} \cdot d \tag{2}$$

This is a variant of eigenvector centrality, and Bonacich (1991[3]) has shown that, for a finite solution to exist, δ must be bounded above by $\frac{1}{\lambda_{max}}$, where λ_{max} is the largest eigenvalue of A.

Proposition 4.1. *λ_{max} is strictly larger than zero for all tournaments where no player only beat players with no wins, or where no player only lost to players with no losses.*

Proof. As A is a real, positive square matrix, the Perron-Frobenius theorem ensures the existence of a unique, real, positive eigenvalue for A, so only the case of $\lambda = 0$ need be checked. Now, A is nilpotent if and only if G is acyclic (that is, contains no directed cycles). But if this were the case, then G must have at least one vertex with no in-coming arcs and one vertex with no out-coming arcs (Ambec & Sprumont, 2002[1]); that is, at least one player must have won no games and at least one player must have lost no games. (Note that this could be the same player in the case of r draws.)

Now consider the subgraph after deleting those players who did not win a game and those players who did not lose a game (i.e. players with no out-coming arcs and players with no in-coming arcs, respectively). Since every subgraph of an acyclic graph is itself acyclic, this subgraph also must contain at least one player with no wins and at least one with no losses. That is, there must have been at least one player who only won against players with no wins, and at least one player who only lost against players with no losses (again, possibly the same player). If this is not the case, then G contains directed cycles, and so $\lambda_{max} > 0$.

□

Corollary 4.2. *The solution for T corresponding to λ_{max} is positive.*

Proof. The statement follows directly from the Perron-Frobenius theorem. □

So T can be calculated from (2) if G is not acylic, but to calculate the likelihood of this is challenging; Robinson (1973[12]) showed that R_n, the number of acyclic digraphs with n labeled vertices, is:

$$R_n = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} 2^{k(n-k)} R_{n-k} \ .$$

Of course, our situation is constrained somewhat by the fact that the maximum in-coming or out-coming arcs a node can have is r. Park and Newman (2005[11]) note that there has never been a season in the American college football league in which there has been no cycles in the directed network; however, the additional edges arising from the fact that draws are not permitted significantly increases the chances of this result, as compared to chess.

It is tempting to try to calculate a more exact probability by assuming that the directed win graph of large chess tournaments could be roughly approximated by a binomial distribution with r rounds and $p = 0.31$ the historical winning ratio from a representative player's perspective. However, as strength is not homogenous, the probability of a win against the representative participant is naturally also variable and is increasing in strength. That said, the Swiss pairing system ensures that players of similar strength are paired against each other after the first round; that is, on average, those who win more games are paired against each other, and likewise those who lose more games are paired similarly. Extending the model to estimate the probability of G being acyclic thus requires an estimation for the distribution of wins in a chess tournament when accounting for the pairing algorithm and heterogeneity in the probability of winning, depending on the opponent. For now, G is assumed to contain directed cycles, and consequently A is not nilpotent.

Within the bounds dictated by $[0, \frac{1}{\lambda_{max}}]$, δ is the sole free parameter for the measure (it makes little intuitive sense to set δ negative, thereby punishing a player for indirect wins). As δ approaches zero from above, the T-score places increasingly less emphasis on indirect wins, to the point where, at $\delta = 0$, the tie-break measure simply equates to the *Won Games* system. The choice of the 'best' δ, however, is to be investigated in the research.

Finally, it should be mentioned that this derivation of T on the basis of a directed 'wins' network could theoretically produce rankings whereby players who scored *less* in their primary score than other participants are ranked higher on the T-measure. Given the stated constraint that the traditional chess scoring system would not be altered, it must be imposed that the proposed measure be *only* used to split ties between players with the same primary score.

5 Empirical Analysis

As mentioned above, a preliminary empirical step should be to analyse chess tournament results under the Swiss pairing system to be able to approximate the distribution of wins for the parameters (n, r). Combining a measure for the distribution with an analysis of the Swiss pairing algorithm will help to determine the probability of A being acyclic; if it is insignificant, then the calculation of the proposed tie-break scores can be applied in events.

The main objective of the empirical research is to analyse data to get the best estimate possible for δ. The T-score must be able to challenge the *Fairness* levels of the current standard measures, particularly Buchholz. One obvious problem is that the degree of *Fairness* will inevitably be subjective to some extent. However, in the first instance, it is sensible to consider cases where the win to draw ratio of tied players is held constant (and so Buchholz should be considered the first-best measure) to ensure that the proposed system computes identical rankings. This process should limit the range in which δ can fall. (Of course, it is possible that the largest eigenvalue is such that no acceptable value of δ can be found within the range.)

As a second step, those tournaments in which the chess public and administration have deemed Buchholz to have failed (such as the 2011 European Championships) should be considered. For the proposed system to be of value, the T-scores must predict rankings 'superior' to those of Buchholz for some δ lying in the newly calibrated range. The definition of 'superior' will again be somewhat subjective, so the focus should be on tournaments in which there has been unanimous agreement by chess commentators and officials on both the flawed Buchholz predictions and the 'true' rankings. Iterations of this kind will narrow the parameter range further.

The strength of the measure's success in addressing the draw issue is less easily verified. One unresolvable issue at the outset is whether decisive results truly give a better reflection of fighting chess and thus should be weighed disproportionately higher than draws at all. It is also arguable as to how effective such a measure in a tie-break system would be in deterring quick draws; a strong player may yet believe that the expected benefits to the *primary* score through a strategy of quick draws outweigh the certain loss in tie-break strength.

Data for all completed tournaments for the past five years is readily available, as in 2008 FIDE began requiring tournament directors to submit complete game results (rather than just a report on a player's overall performance in the tournament). This is available from the FIDE website[8], and also from Chess Results[9], an online server for tournament data. Furthermore, results for many completed tournaments before 2008 are also available, as many organisers chose to report the complete results voluntarily. On average, several dozen tournaments are completed and reported each week.

Open tournaments vary in size from 20 to several hundred participants, with the mode being roughly 40. It is possible that different values of δ could be appropriate to different participation sizes, and that a suggested partitioned implementation of "δ for n up to 100, and $\alpha\delta$ for $n > 100$" is optimal (with $0 < \alpha < 1$, to focus more on direct wins as the network grows). Fortunately, the data pool is large enough that the iterative process for calibration outlined above can be repeated for different ranges of n, and these size-dependent values of δ can be compared.

6 Summary

The primary purpose of the research is to address the two predominant issues in international chess at present: unreliability of current tie-break systems, and the increasing proportion of quick draws among strong players. The secondary purpose is to investigate the properties of the directed networks created by player wins in chess tournaments.

These aims will be dually achieved through the development of a new tie-break system based on a measure from directed networks. Using the results from eigenvector centrality, participants receive a tie-break score based on a weighted sum of direct and indirect wins through the network. The weighing parameter should be calibrated from data of past completed tournaments to ensure the measure is fair and robust.

If the new tie-break system is proved empirically robust, its contribution would be firstly to reduce 'unfair' tie-break decisions (as illustrated at the outset), and secondly, to offer an incentive for players to play more combatively. An extension offering a further advantage would be that the measure could be applied uni-

[8] `www.FIDE.com`

[9] `http://chess-results.com/`

formly across both Swiss and round-robin system events. As mentioned above, separate tie-break systems are employed for all-play-all tournaments, as the Buchholz scores for tied players are necessarily identical. However, in the (reasonably likely) event of the *Direct Encounter* between two (or more) tied players being even, the proposed measure offers a solution. In this instance where opposition strength is homogenous, tied players would be split on the basis of a combination of personal combativeness (through number of direct wins) and the discounted, *weighted* sum of opponents' scores, where wins over the more combative and (on average) higher performing opponents are given more credit. Thus, the tie-break system could be universally applied across the two most common chess tournament formats.

References

[1] S. Ambec and Y. Sprumont. Sharing a river. *Journal of Economic Theory*, 107:453–462., 2002.

[2] D.R. Appleton. May the best man win? *Journal of the Royal Statistical Society. Series D (The Statistician)*, 44(4):pp. 529–538, 1995.

[3] P. Bonacich. Simultaneous group and individual centralities. *Social Networks*, 13:155–168, 1991.

[4] K.A. Butler. Some problems in paired comparisons and goodness of fit for logistic regressions. *PhD Thesis, Simon Fraser University*, 1997.

[5] L.C. Freeman, S.P. Borgatti, and D.R. White. Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, 13:141–154, 1991.

[6] M.O. Jackson. *Social and Economic Networks.* Princeton University Press, Princeton, NJ, USA, 2008.

[7] R. Kasimdzhanov. Open letter with a proposal. ChessBase News, `http://www.chessbase.com/newsdetail.asp?newsid=7387`, July 2011.

[8] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18:39–43, 1953.

[9] I. Leong and L. Weiwen. A cure for sad (severe acute drawitis). ChessBase News, `http://www.chessbase.com/newsdetail.asp?newsid=2719`, November 2005.

[10] J. Nunn. The draw problem: a simple solution. ChessBase News, `http://www.chessbase.com/newsdetail.asp?newsid=2729`, November 2005.

[11] J. Park and M. Newman. A network-based ranking system for us college football. *Journal of Statistical Mechanics: Theory and Experiment*, (P10014), 2005.

[12] R.W. Robinson. Counting labeled acyclic digraphs. In F. Harary, editor, *New Directions in the Theory of Graphs*, pages 239–273. Academic Press, NY, 1973.

[13] J. Sonas. What exactly is the problem of draws in chess? ChessBase News, `http://www.chessbase.com/newsdetail.asp?newsid=7499`, September 2011.