

Computing power structures in directed biosocial networks: flow percolation and imputed conductance

Kevin Fujii ^{*}, Hsieh Fushing [†]

Dept. of Statistics, UC Davis,

Brianne Beisner, and Brenda McCowan

Department of Population Health & Reproduction, UC Davis,

California National Primate Research Center

Abstract

We introduce a new perspective on power structure to describe the global pattern of dominance among subjects in a competitive society. Based on pairwise conflict data, we compute and reveal society’s power structure as a composite hierarchy, which is often overlooked when the somehow elusive linear ranking is focused. The computing of such a power structure involves combinatorial optimizations and empirical dominance transitivity based on directed conflict connectivity. The first biological implication of a power structure is its manifestation of dominance flows, which fundamentally constitute a foundation underlying the dynamics under study; that is, other systemic aspects or behaviors need to conform with such a structure as an embedded skeleton. The second implication is its degrees of sensitivity to perturbations, which critically determine the size of its bootstrapping ensemble and the room allowing for modeling. Two computational devices, flow percolation and imputed conductance, are developed to sharpen estimations of dominance probability matrices. We analyze two real datasets of decisive conflict outcomes, the first involving college football teams, and the second involving an adult rhesus macaque society in captivity.

Introduction

In a competitive society of individuals or teams, it is often of interest to infer the architecture within the society’s relational complexity revealed by the observed competitions. Most commonly, such competitions are pairwise or dyadic between the subjects in the society. We emphasize the importance of computing this architectural complexity and call it the power structure of the society, rather than inferring just the ranking order of the subjects. The *power structure* of a society takes the form of one single or several partially connected hierarchies (Simon, 1962), in which any individual node is positioned with clearly expressed local as well as global flows of outward or inward “dominance” relationships. For expositional convenience, we use “dominance” as the generic directed relation throughout this paper. The power structure of a society works to transcend the typical confines of a ranking. In contrast, the desire to rank the subjects habitually refers

^{*}Supported by NSF grant: DMS-1007219(CDI program), awarded to Hsieh Fushing.

[†]Author for correspondence: fhsieh@ucdavis.edu.

to linearity of dominance. However, such a linear ranking may not be appropriate for a particular society, or it may not even be inferable from the observed conflicts. Instead, we aim to construct a power structure of the subjects in which we preserve certain local features of the linear ranking, but acknowledge situations in which direct comparisons between subjects likely are unfounded.

Power structure is a non-parametric global construct. Construction of the power structure requires not modeling but rather computing and organizing substructures in the network. We develop computational algorithms to achieve this constructing task. Several innovative computational devices are developed here: percolation, transitivity estimation, conductance imputation, and hierarchical network organization. These devices are put together to overcome the difficulties encountered in Fushing et al. (2011) via the inclusion of only low-order dominance paths and the conservative nature of the dominance odds computation.

As in general settings, one may think of the collection of competitions as a weighted, directed graph of pairwise conflicts. Observed data is represented by a conflict matrix $C = [c_{ij}]$, where c_{ij} is the number of competitions in which subject i dominates subject j . Our percolation algorithm relies on the transitivity of dominance information to impute dominance probabilities between subjects for which no direct conflicts are observed (i.e., $c_{ij} + c_{ji} = 0$). That is, if we observe the dominance of subject $i_0 = i$ over i_1 , i_1 over i_2 , ..., and i_m over $i_{m+1} = j$, we can infer the dominance, to some degree, of subject i over subject j . Such a chain of dominance information is referred to as a *dominance path*. In order to take advantage of dominance transitivity, we must be able to identify these dominance paths in the graph of conflicts. We provide R code in Appendix D that will find all dominance paths of any specific order. By identifying dominance paths in the conflict network, we can determine all pairs of subjects that have any amount of dominance information transmitted between them. If a pair of subjects $\{i, j\}$ has no dominance paths from one to the other, then we cannot reasonably infer any dominance information about a hypothetical conflict between the two subjects.

We are not aware of previous works that attempt to infer an explicit hierarchical power structure of a society, given pairwise conflict information. An early attempt is a corporative ranking hierarchy for a rhesus macaque monkey society in Fushing et al. (2010). There exist several algorithms to compute a linear ranking from such information. The Bradley-Terry model (Bradley and Terry, 1952) is commonly used with pairwise comparison data and involves the assignment of a dominance index to each subject; a likelihood function is obtained, then maximized with respect to the dominance indices. Thus, the linear ranking is brought out by permuting the estimated dominance indices. Elo ratings (Elo, 1978) similarly assign a numerical rating to each subject, and are popular for their ease of recalculation upon the acquisition of additional conflict data. These models, though, use only direct information from the outcomes of the conflicts and ignore indirect dominance information transmitted transitively through the conflict network. Fundamentally, the models' linearity assumptions are inappropriate and too restrictive when making inferences about the power structure of a competitive society.

In contrast, our percolation algorithm makes use of dominance paths to impute dominance probabilities, taking advantage of the transitivity of dominance information for inference of power structure within a network. We further suggest a conductance approach for improving the efficiency of this percolation algorithm by directly utilizing dominance paths within the network. This conductance algorithm is applied to two data sets of interest. Inferences about power structure of each network are reported and compared to existing methods.

We further discuss the implications of power structure, as providing a foundation for reasoning why bootstrapping on directed graphs in general may be difficult or even impossible. Consequently, this result implies that modeling and hypothesis testing on directed networks without extracting power structures are

typically not feasible. We demonstrate through a real case that the power structure has to be taken as a deterministic frame, to which violation is not permissible. That is, it has to be a frame of all other aspects of dynamics under study. Also, from a behavioral ecology perspective, this power structure, constructed via a key behavior would likely become the skeleton of our understanding of the behavioral dynamics of a particular society. Since a behavior is a facet of such dynamics of interest, and typically there are many behaviors of the same dynamics that are considered and observed, such a skeleton could provide a unified framework to holistically understand all other behaviors and interactions as all facets of dynamics as one whole.

Determining transitivity and prior distributions

Fushing et al. (2011) incorporated indirect dominance information by introducing the concept of dominance paths. The probability of dominance p_{ij} for subject i over subject j is drawn from the posterior Beta distribution $Beta(c_{ij} + 1, c_{ji} + 1)$ for every subject pair $\{i, j\}$ for which $c_{ij} + c_{ji} > 0$. The randomness of the assignment of these dominance probabilities reflects the uncertainty in conflict outcomes. Dominance paths with high component dominance probabilities are included in odds calculations as indirect conflict information. It is almost always the case that, with pairwise conflict data, there exists a subject pair $\{i, j\}$ such that no direct information is observed (i.e., $c_{ij} + c_{ji} = 0$). Thus, we must rely on indirect information, quantified by dominance paths, to help impute dominance probabilities. However, the thresholding involved in selecting dominance paths for odds calculations by Fushing et al. (2011) is rather conservative; it disqualifies most dominance paths of moderate order and even some of low order that contain valuable dominance information.

To incorporate more dominance information, we introduce a trickling-down percolation algorithm to infer the hierarchical power structure within a network in the next section. This algorithm takes advantage of information transitivity to impute dominance information between any two particular nodes linked by dominance paths. Here we also impute missing dominance probabilities based on the Beta random field approach introduced by Fushing et al. (2011). However, we endeavor to make more efficient uses of dominance paths of many orders, and to extract the empirical transitivity in order to facilitate better prior Beta distribution from the data-driven perspective.

Let there be N subjects in a society, each of which is involved in at least one conflict. We denote a particular dominance path from subject i_0 to i_{m+1} passing through subjects i_1, \dots, i_m by $\sigma(i_0, i_1, \dots, i_m, i_{m+1})$. The *order* of this dominance path is defined as the number of intermediate nodes in the path. Thus, the dominance path $\sigma(i_0, i_1, \dots, i_m, i_{m+1})$ is of order m . Since cyclical dominance paths in the network of conflicts contain no dominance information, we restrict dominance paths of order m to consist of exactly $m + 2$ distinct nodes. We enumerate dominance paths of order m by $\sigma_1^m, \dots, \sigma_{K_m}^m$, where K_m is the total number of dominance paths of order m . We index the f^{th} element of dominance path σ_k^m as $\sigma_k^m(f)$, where the starting node corresponds to $f = 0$ and the ending node corresponds to $f = m + 1$.

Let $D_m = [d_{ij}^m]$ be the order m dominance path matrix, where d_{ij}^m is the number of dominance paths of order m beginning at subject i and ending at subject j . We note that C^{m-1} gives an element-wise upper-bound for the entries of D_m , as C^{m-1} enumerates dominance paths with cycles allowed. Furthermore, for simplicity of notation, let $w_i = \sum_j c_{ij}$ be the number of wins by subject i , or the number of times subject i was observed to dominate another subject. Conversely, let $l_j = \sum_i c_{ij}$ be the number of losses by subject j , or the number of times subject j was observed to be dominated by another subject.

A Beta distribution that is a generalization of that which was used in Fushing et al. (2011), and allows context-dependent flexibility, is proposed as follows. After observing the direct conflict data c_{ij} and c_{ji} , $Beta(\alpha c_{ij} + \beta, c_{ji} + \beta)$ is the posterior distribution for the dominance probability of i over j using a prior distribution of $Beta(\beta, \beta)$ with each dominance action worth α units. Methods for strategically choosing α and β are described below.

For values of α and β used in the posterior Beta distribution to be context-dependent, they need to reflect the influence of an observed dominance action on the underlying dominance probability. An expert in the appropriate field of study may have the context necessary to accurately estimate α and β intuitively. In the absence of such an authority, though, α and β are better estimated through a data-driven means.

Nikolić (2007) discusses two characteristics of directed networks: transitivity and additivity. A network is *transitive* if the directions of dominance paths are consistent with observed dominance actions; the network is *additive* if the magnitudes of the directed edges are preserved in a dominance path from the starting subject to the ending subject. As additivity is a stronger condition than transitivity and is rarely observed in pairwise conflict networks, we focus on using transitivity to infer values of α and β .

Nikolić’s method for measuring network transitivity involves inspecting all $\binom{N}{3}$ groups of three nodes (called *triples*), and, if edges between all three nodes exist, recording whether those edges are transitive rather than cyclic. If there is a large proportion of transitive triples, we may infer that dominance paths of order 1 carry strong dominance information from the starting subject to the ending subject. On the other hand, if there is a relatively large number of cyclic triples, then dominance paths of order 1 are not necessarily reliable indicators of dominance of the starting subject over the ending subject. We calculate *order-1 transitivity* as the proportion of transitive triples among all observed triples. Triples without three observed edges are ignored in this calculation.

We can extend this idea to help estimate α and β just by observing these transitive or cyclic triples. Suppose there is a dominance path from subject i_0 to subject i_2 , passing through subject i_1 . If the network is highly transitive, then this dominance path strongly suggests that a conflict between i_0 and i_2 would end in a victory for i_2 . High transitivity in a network also implies that there is not much variability in dominance action outcomes. That is, there are not many “upset” victories that would create cyclic triples. In highly transitive networks, then, we can infer that each dominance action is a very reliable indicator of true dominance. That is, observing subject i dominating subject j in a highly transitive network is indicative of a dominance probability $p_{ij} \approx 1$. On the other hand, if the network is not transitive, an observed dominance action of subject i over subject j may not affect the dominance probability p_{ij} by a significant amount.

Transitivity, then, can be used to help determine appropriate values of α and β for use in the posterior Beta distribution in Equation (1). Transitivity estimates the probability that a length-1 dominance path is representative of actual dominance. Thus, the probability that an order-1 dominance path is traversed in the percolation algorithm should be approximately equal to transitivity. In order to determine α and β , we equate the transitivity of the network with the squared mean of the posterior Beta distribution. That is, we choose α and β such that $\left(\frac{\alpha+\beta}{\alpha+2\beta}\right)^2$ is approximately equal to the transitivity. With the common choice of $\beta = 1$, we choose to set $\alpha = \frac{2\sqrt{T_1}-1}{1-\sqrt{T_1}}$, where T_1 is the order-1 transitivity of the observed network.

In the two data sets described in detail later in this paper, we observe different levels of transitivity. In the 2011 NCAA football data, there is a non-negligible proportion of non-transitive triples. The order-1 transitivity of the NCAA football data is 0.7194, meaning that only 71.94% of observed triples were transitive. Therefore, dominance paths of order 1 should not always be considered indicative of the dominance of the starting subject over the ending subject. We may want to modestly choose values of α and β to reflect some

degree of uncertainty in dominance probabilities, even after observing all 677 games in the 2011 season. The choice of $\alpha = 4.6$ and $\beta = 1$ gives us the posterior distribution

$$p_{ij}|c_{ij} \sim \text{Beta}(5.6, 1)$$

when we observe team i defeating team j once in the 2011 season. This distribution has a mean of 0.72 and a standard deviation of 0.13, which seem to accurately reflect our uncertainty in dominance probabilities as expressed through order-1 transitivity.

In the monkey “silent bared-teeth display” (SBT) data, though, we calculated order-1 transitivity of 1.0000, so every observed triple was transitive. It follows that every observed SBT action was consistent with some underlying hierarchical social structure. Therefore, dominance paths surely carry the same amount of dominance information as a directly observed SBT action. Since we are relatively certain that SBT actions almost always indicate true dominance, we can assign values of α and β such that the posterior Beta distribution of each p_{ij} is highly skewed towards 1, whenever we observe i dominating j . The network is perfectly transitive, so we choose α and β such that the mean of the posterior Beta distribution is nearly 1. A choice of $\alpha = 18$ and $\beta = 1$ leads to a posterior distribution of

$$p_{ij}|c_{ij} \sim \text{Beta}(19, 1)$$

if we observe a single SBT action in favor of i over j . The posterior distribution of p_{ij} here has a mean of 0.95 and a standard deviation of 0.05, which represents relative certainty of true dominance, given an observed SBT action. The parameters α and β can of course be modified to achieve another desired posterior Beta distribution.

We also note that higher orders of transitivity may be used, simply by counting larger cycles instead of non-transitive triples. However, this information may be redundant, as we observe that these higher-order transivities are largely dependent on that of order 1.

The Percolation Algorithm

The percolation algorithm is originally conceived with the aim of making more efficient use of dominance paths of many orders. It is defined as follows. We select a starting node i_0 such that $w_{i_0} > 0$, and a candidate neighbor i_1 uniformly at random such that $c_{i_0 i_1} > 0$. Then, we generate a Bernoulli random variable X_0 with probability of success drawn from the distribution

$$\text{Beta}(\alpha c_{i_0 i_1} + \beta, \alpha c_{i_1 i_0} + \beta), \tag{1}$$

where $\alpha \geq 1$ and $\beta \geq 1$ are chosen to reflect the influence of an observed dominance action on the dominance probability.

If the simulated Bernoulli random variable X_0 is equal to 1, then we repeat the process outlined above, treating i_1 as the starting node. If $X_0 = 0$, then we terminate the process. Upon the termination of this process, we are left with a path of “accepted” nodes (that is, all but the final candidate node, corresponding to $X_m = 0$). Suppose that the path of accepted nodes is i_0, i_1, \dots, i_m , and the final candidate node is j . Then, we assign a simulated dominance action for i_a over i_b for any $a, b \in \{0, \dots, m\}$ such that $a < b$. Finally, we assign a simulated dominance action for j over i_m . We repeat this process starting at each node many

times, and record the simulated dominance actions in an ensemble matrix E , whose entries e_{ij} enumerate the number of dominance actions for subject i over subject j were simulated by the aforementioned percolation algorithm. We finally convert the ensemble matrix to a dominance probability matrix P , where the entries p_{ij} are calculated by

$$p_{ij} = \frac{\alpha e_{ij} + \beta}{\alpha(e_{ij} + e_{ji}) + 2\beta}, \quad (2)$$

the posterior mean of the previously utilized Beta distribution, treating the simulated dominance actions as observed. We use simulated annealing (Kirkpatrick et al., 1983) to reorder the rows and columns of P such that P is as close as possible to upper-triangular. Let ρ be a permutation $(1, 2, \dots, N)$ that defines a reordering of the rows and columns of P , and let $P[\rho]$ be the reordered version of P according to ρ . We use the cost function

$$Cost(P[\rho]) = \sum_{i=2}^N \sum_{j=1}^{i-1} \max(0, -\log[2(1 - p_{\rho,ij})]) e^{\frac{(N+1-j)(i-j)}{N^2}}, \quad (3)$$

as in Fushing et al. (2011), where $p_{\rho,ij}$ is the (i, j) th element of the reordered matrix $P[\rho]$. The optimal reordering $\rho^* = \arg \min_{\rho} Cost(P[\rho])$ of the rows and columns provides the ranking of the subjects.

While the percolation algorithm allows us to impute dominance probabilities by exploring dominance paths, we have no guarantee that every dominance path (of a reasonable order) will be explored. Therefore, we may be able to more effectively utilize transitivity of dominance information by identifying all dominance paths up to a particular order. We use these dominance paths directly in the following conductance method for obtaining a hierarchical power structure.

Conductance for computing power structure

The potential disadvantage to our percolation algorithm, mentioned above, is that we have no guarantee that every dominance path (of a reasonable order) will be explored. In order to incorporate all dominance paths (up to a particular order) we need to effectively use the transitivity of dominance information. With this idea in mind, we modify our percolation method to include all dominance paths up to a certain order. This method, called conductance method, weights each dominance path's contribution to the imputed conflict matrix (or ensemble matrix) by its probability of being successfully traversed in the percolation algorithm. This way of imputing missing dominance information resembles a non-stochastic version of the percolation algorithm. However, we note that the method can be adjusted to remain stochastic, in which case we may use it to extract further information regarding the distribution of dominance probabilities or rankings. A full description of the necessary calculations is available in Appendix A.

By identifying the dominance paths up to a particular order, we can ensure that all dominance information, direct or indirect, implied by the observed conflicts is taken into account when estimating dominance probabilities. We use this conductance method as a replacement for performing a large number of iterations in the previously described percolation algorithm. Both the percolation and conductance algorithms simulate the observed conflicts, after using them to compute preliminary estimates of certain dominance probabilities. After the algorithm (either percolation or conductance) is completed, we use the indirect information obtained through dominance paths to update existing dominance probabilities and estimate those that have not yet been calculated.

Via the calculations in Appendix A, we collect all direct and indirect information into E^* , the $N \times N$ non-stochastic ensemble matrix. E^* contains a combination of directly observed conflicts and imputed information from dominance paths. To obtain a permutation that provides some information regarding the ranking of subjects, but is not necessarily a linear ranking, we use the simulated annealing algorithm described in the previous section with the cost function displayed in (3). The dominance information from the conductance algorithm produces a power structure - it is information regarding the relative rankings of subjects that is not necessarily structured as a linear ranking.

Simulation Study

To evaluate the effectiveness of using dominance paths in the conductance algorithm to detect power structure in a competitive network, we simulate networks as follows. We assign each of 25 subjects a unique dominance index d_i between -5 and 5. The probability that subject i dominates subject j in competition is assigned by:

$$p_{ij} = \frac{1}{1 + e^{0.5(d_j - d_i)}},$$

so that the relationships between the subjects are linear. That is, if $d_i > d_j > d_k$, then $p_{ik} > 0.5$.

The dominance indices were chosen such that five groups of five subjects each are discernable, and the dominance indices of the subjects within any particular group are similar. As a result, the dominance probability matrix has a block-like structure, as shown in Figure (1), on right.

We create a fixed “schedule” of 125 conflicts such that each subject participates in exactly 10 conflicts, competing against the same 10 opponents each time the simulation is run. We then use conductance to estimate the dominance probability matrix and obtain a ranking. Transitivity is estimated to be 0.7778, indicating a moderate level of information that can be inferred transitively via dominance paths. Such a level of transitivity is reasonable, given the structure of the simulated data. Dominance paths between tiers of subjects are highly informative, while paths within tiers are of lower importance. The average estimated dominance probability matrix from 100 simulations is also shown in Figure (1), on left.

While the estimated dominance probability matrices for each individual simulation did reflect the tiered structure of the underlying competitive society, it is clear that this is captured on average by our algorithm.

The conductance algorithm used to estimate the dominance probability matrix causes us to incur a systematic bias, shown in Figure (1), at bottom. Specifically, it underestimates dominance probabilities that are greater than 0.5, and as a result overestimates those probabilities less than 0.5. Because of the uninformative prior distribution placed on the dominance probabilities, the posterior estimates are shrunk towards 0.5. We note that when large amounts of data are available, the bias disappears due to the design of the conductance algorithm. However, with realistic amounts of data, the bias remains.

Bias in the estimation of dominance probabilities may arise in two forms. The estimated probability could be smaller than 0.5 when the true value is larger than 0.5, or vice versa. Alternatively, the estimated probability could be on the correct side of 0.5, but closer to 0.5 than the true value actually is. The former is likely to be due to uncertainty in conflict outcomes, especially in the presence of upset victories, and therefore overvaluing certain observed dominance actions. The latter is generated from conservativeness of estimates as a result of a prior distribution centered at 0.5. The amount incurred of either type of bias depends largely on the choice of α and β values used to update the Beta distributions that model the dominance probabilities, and we hope to minimize total bias, or at least achieve a balance between these two sources of bias.

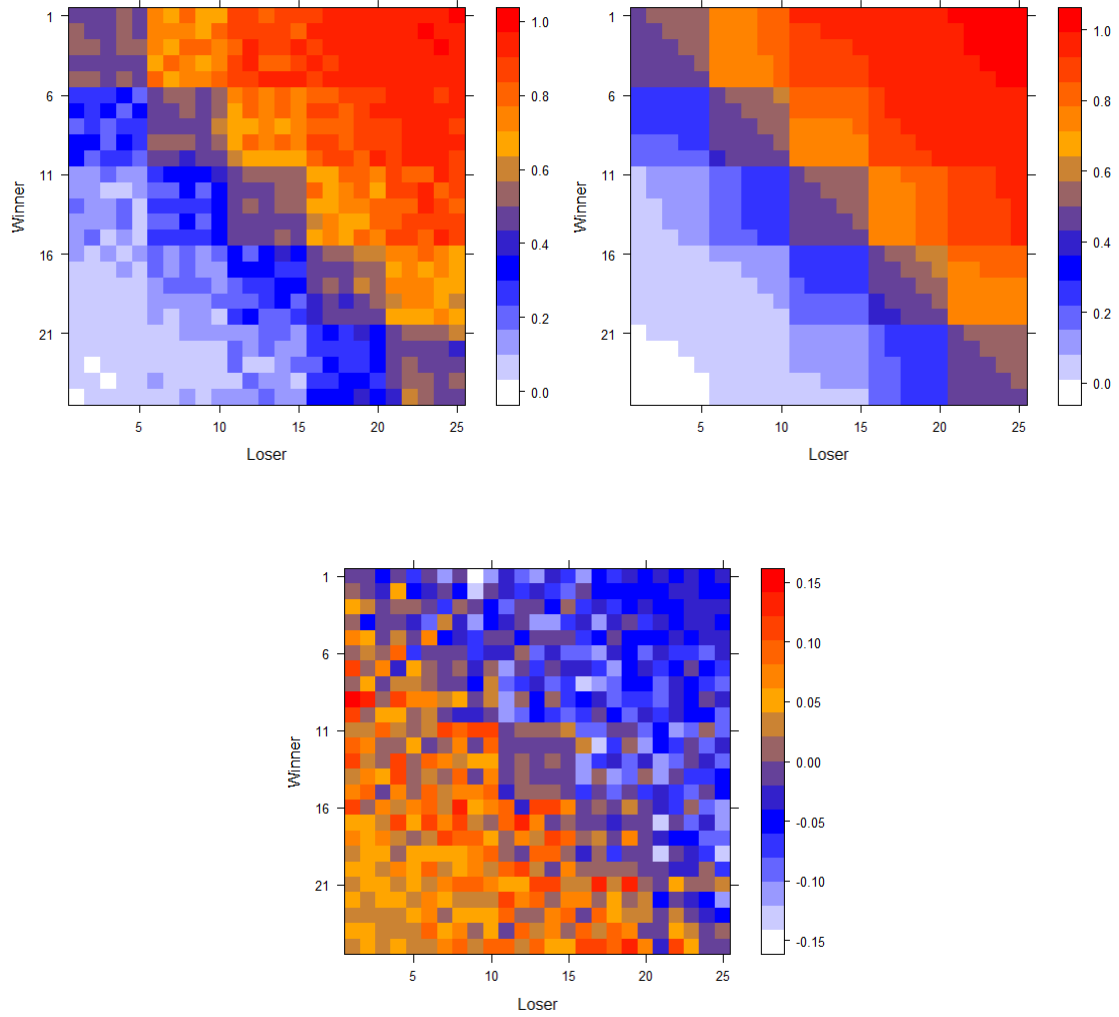


Figure 1: The average estimated dominance probability matrix from 100 sets of simulated data (left), and the underlying dominance probability matrix (right) used to generate said simulated data. The color represents the dominance probability of the row subject over the column subject. Estimation of dominance probabilities tends to be biased towards 0.5. The difference between the estimated dominance probability matrix and the true dominance probability matrix is shown at bottom, displaying the systematic bias.

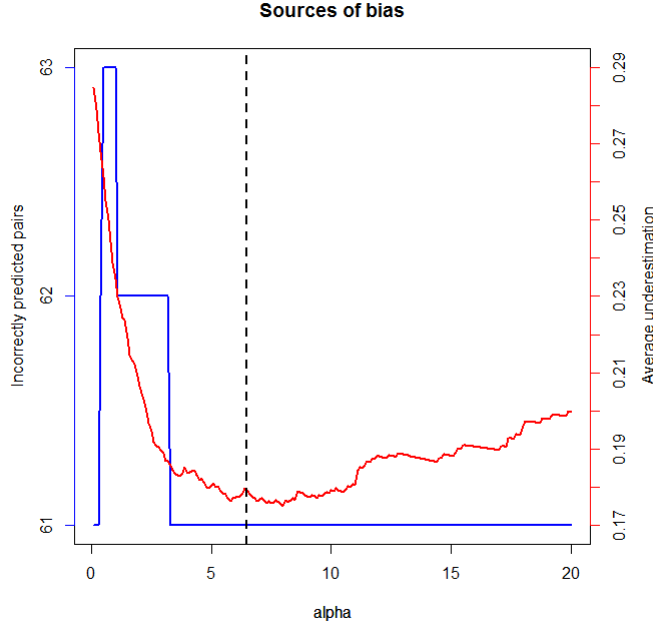


Figure 2: Dominance probabilities estimated on the wrong side of 0.5 (blue) and average underestimation of dominance probabilities greater than 0.5 (red). The optimal value of α for this simulation is displayed as a dotted vertical line, which approximately minimizes both sources of bias.

To see how these biases may arise, we return to the previously described simulated data. With a transitivity of 0.7778, we estimate α to be approximately 6.468 with $\beta = 1$. In Figure (2), values of α between 0 and 20 are tested, and the estimated dominance probabilities are compared to the true values shown in Figure(1). We find that the value of α that approximately minimizes both the number of dominance probabilities estimated on the wrong side of 0.5 (that is, the number of pairwise dominances that our method incorrectly infers) and the average underestimation of dominance probabilities greater than 0.5 (i.e., shrinkage towards 0.5) is very close to 6.468. This indicates that the transitivity of a directed network tells us a great deal about how to value indirect information, and is quite useful when deciding how update the Beta distributions. We note that while the magnitude of the shrinkage towards 0.5 seems large in this simulation, there is at most one direct interaction between any given pair of subjects. With such little conflict data, it is unreasonable to expect to estimate dominance probabilities even remotely close to 1 or 0.

Data descriptions

The first of the two sets of data to be used in this paper is taken from a society of monkeys at the California National Primate Research Center (CNPRC). The society of $N = 84$ adult rhesus macaques express peaceful submission via an action called the “silent bared-teeth display” (SBT) in which the lips are pulled back to bare the teeth, described in Beisner et al. submitted. These peaceful submission events are taken as conflict data because the monkey giving the SBT is subordinate to the monkey receiving the SBT. Previous studies of macaques indicate that SBTs are very reliable communications of dominance relationships because these signals are almost entirely unidirectional within a given dyad (de Waal and Luttrell 1985; Preuschoft 1995;

Flack and de Waal 2007). SBTs are therefore ideal for constructing the power structure of the society. The rhesus macaque social group was observed from May to December 2009 for a total of 182 hours (6 hrs/day, 4 days/week on a rotating schedule; see Beisner et al 2011). Conflict data, including SBTs, were recorded using an event sampling design (see McCowan et al 2011 for further detail). The rhesus macaque conflict data consists of 457 observed peaceful SBT signals, constituting 369 unique weighted edges. The monkeys are allowed to freely interact within the group, so some pairs of monkeys will interact several times, while others will not encounter each other in conflict.

The second set of data to be analyzed here is the regular season results from the 2011 National Collegiate Athletic Association (NCAA) football season. College football teams from $N = 120$ universities in the Football Bowl Subdivision (FBS) of the NCAA compete amongst each other, and are ranked by computer algorithms and expert polls. These rankings are compiled into one overall ranking called the Bowl Championship Series (BCS) ranking. BCS rankings at the end of the season are used to determine which two teams will compete in the National Championship game, and which other teams are eligible in sponsored and heavily publicized bowl games. A total of 677 games between FBS teams were played in the 2011 regular season. Because of the rigid schedule set by the NCAA, these conflicts result in 676 unique directed edges, with all edges but one having a weight of 1. FBS teams are divided into 11 conferences of 8-13 teams each, and by design, most conflicts occur within a conference. Thus, much imputed dominance information will depend on the results of inter-conference games.

We will apply the previously described conductance algorithm to analyze these two directed networks. While both networks consist of competitive actions, we find that the structures contrast greatly. In terms of determining who participates in conflicts, the NCAA schedules between 10 and 13 games for each team competing in the Football Bowl Subdivision (FBS), the top level of college football. Most of these games occur within the team's conference, while the remaining few games are played between conferences. The result is a graph with several clearly defined, well-connected conferences. The inter-conference games help to connect the conferences to each other, which in turn define numerous dominance paths between teams in different conferences that otherwise would not have been connected. These inter-conference games, therefore, become quite important when imputing dominance probabilities for pairs of teams in different conferences, as they make indirect comparisons possible. In the monkey SBT data, however, such a schedule is non-existent. The monkeys themselves decide with whom they interact, and to whom they display bared-teeth.

The two networks, as previously mentioned, have substantially different transivities. Because the monkey SBT network is perfectly transitive, each dominance path (of any length) is essentially an indicator of true dominance of one monkey over another. This particular network structure makes it extremely easy to compare monkeys with at least one dominance path between them. There are many pairs of monkeys that we cannot compare, due to the fact that they do not have any dominance paths between them. In contrast, there are dominance paths in both directions between almost all pairs of NCAA football teams, and as a result, the NCAA network is only moderately transitive. While a dominance path from one football team to another carries some dominance information, it is also likely that dominance paths exist in the opposite direction.

Due to the differences in network structures, we take slightly different approaches when analyzing the two societies. With the NCAA football network, we are able to compare nearly any two teams using either directly observed conflicts or dominance paths. As a result, the dominance structure tends to be somewhat linear, except in cases where there is no dominance information transmitted between two teams, or when the information transmitted in both directions is approximately equal. It is these cases in which we are

particularly interested, as they provide information transcending a linear ranking of teams. In particular, we can determine which subsets of teams are roughly of the same status in the network, and thus should not be ranked against each other. We can still provide a linear ranking of football teams, but it would be desirable to have some additional perspective about the teams' relative standing.

In the monkey SBT network, we can easily compare any pair of monkeys with a dominance path of any length between them. We cannot obtain a full linear ranking of monkeys, however, because a substantial number of monkey pairs are not comparable by dominance paths. We will be able to construct a hierarchy based on the partial rankings provided by every dominance path in the network. Such a hierarchy can tell us which monkeys have influence over which other monkeys, and which pairs of monkeys are not comparable.

Results: NCAA Football

We apply the previously described conductance algorithm to the results of the 2011 NCAA college football season. For computational and practical reasons, we identify all dominance paths in the NCAA football network up to order $M = 7$ for use in the algorithm.

The objective of the BCS ranking system is to obtain a linear ordering of all 120 NCAA football teams. As mentioned in previous sections, however, a complete linear ranking of subjects may not be appropriate or inferable from available dominance information. If there exist no dominance paths between two particular teams, we cannot reasonably infer that one team is more likely to defeat the other. Dominance information is nonexistent in this case, directly and indirectly.

Alternatively, we aim to construct a hierarchy of teams. We recall that the cost function via Equation (3) is essential for obtaining an optimal linear ranking. Note that if there is no dominance information between two adjacent teams in the ranking, then the two teams' positions in the rankings may be swapped without incurring additional cost. As a corollary, two teams with nearly equal dominance information between them may be swapped with a minimal increase in cost. Such a small difference in costs may be negligible, after considering the variability in conflict outcomes.

Taking this into account, we would like to determine which pairs of teams may be swapped with minimal increase in cost. Figure (4) displays a matrix of costs after pairwise swaps are made to the optimal linear ranking. Such plots enable us to visualize the tiered nature of the football team hierarchy, as well as the instability in some of the teams' rankings. Either matrix shown in Figure 4 may be used as a dissimilarity or similarity matrix, respectively, within a clustering algorithm to determine the tiers of teams. The resulting clusters, as shown in the hierarchical clustering tree in Figure (5), may be interpreted as tiers. A desired number of tiers can be obtained by cutting the branches of the tree at an appropriate height.

Figure (5) shows the clear division of the top 25 NCAA football teams into two tiers: those ranked 1-12, and those ranked 13-25. Within those tiers, further subdivision of groups of teams is possible, depending on how many tiers are desired. We can see that among the top 12 teams, LSU and Alabama are far distinguished from the others, as none of the other 118 teams had any conductance whatsoever to either team. Arkansas and Oklahoma State seem to be the clear third- and fourth-ranked teams, followed by Kansas State, Stanford, and Baylor. The remaining five teams in the top 12 appear to be very close in ranking, as permuting any of these teams in the optimized ranking results in very small changes in the overall cost.

We find that the construction of this tree, the identification of clusters or tiers, and the calculation of the matrices in 4 provide much valuable information beyond a linear ranking. These figures allow inference about the stability of rankings and helps to identify those pairs of teams that are not comparable with dominance

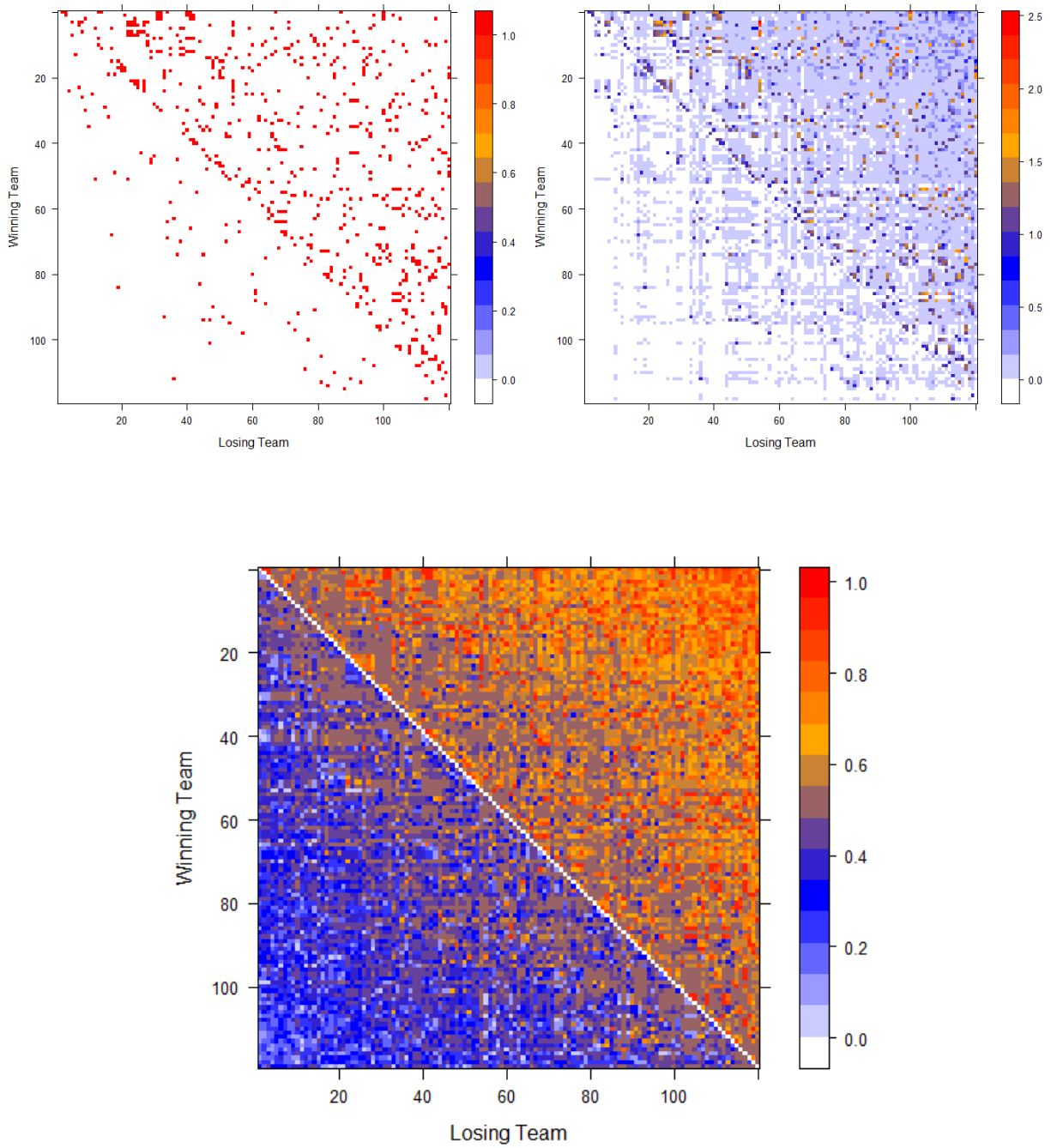


Figure 3: The reordered NCAA football conflict matrix (top left), imputed conflict matrix from Method 1 (top right), and estimated dominance probability matrix (bottom).

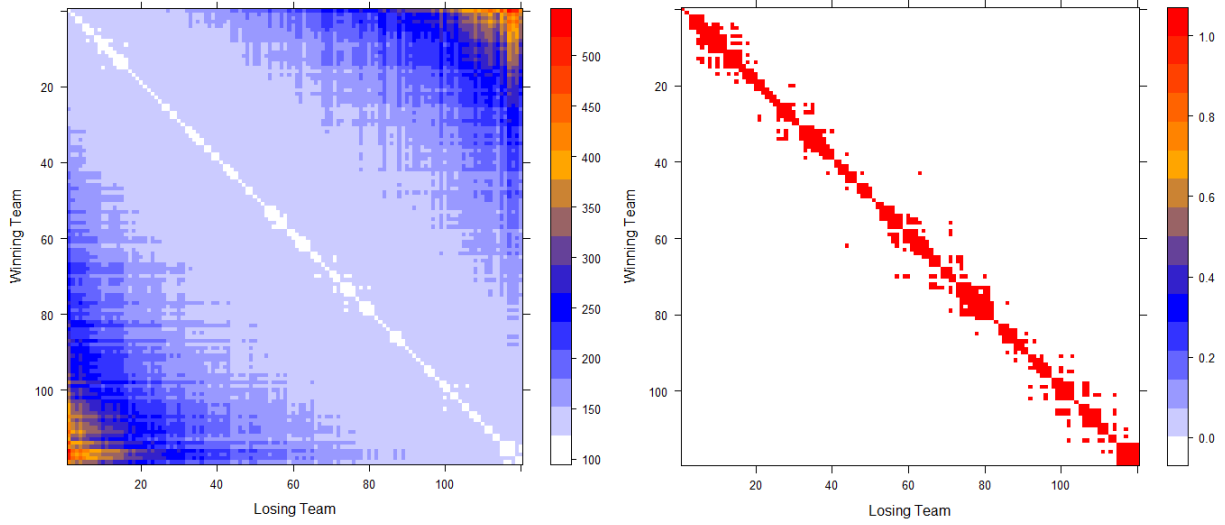


Figure 4: On left, a matrix of costs as calculated by the cost function in Equation (3), the (i, j) th entry of which corresponds to the cost of the optimal linear ranking with the i and j th entries swapped. Costs that have increased by fewer than 1 unit are shown on right. The above matrices indicate that small, local permutations in the optimal linear ranking can be made without significantly increasing the cost. The figure on right helps to display some groups of teams that may be permuted with each other incurring only small increases in cost.

information. The construction of the matrices in (4) also allow us to further refine the final linear ranking from simulated annealing, in case the simulated annealing algorithm does not fully converge to the truly optimal linear ranking.

Results: Monkey Data

Using the conductance algorithm, we can successfully rank the monkeys such that the simulated annealing cost function in Equation (3) is zero. That is, estimated dominance probabilities p_{ij} are less than or equal to 0.5 whenever i is ranked higher than j . However, we must note that there are several possible rankings that also achieve zero cost, and in fact, many of the monkeys were never dominated by another in an SBT conflict. Those monkeys are interchangeable in the rankings, with respect to the cost function in Equation (3). The implication is, by restricting inference to linear rankings, we are forced to disregard any underlying hierarchical SBT power structure in the society of monkeys. Inspection of the estimated dominance probability matrix indicates that there exist many pairs of monkeys that are not comparable by dominance paths. That is, as opposed to the NCAA football data, we do not observe any dominance paths between most pairs of monkeys, which makes it impossible to impute SBT dominance information. The cost function in Equation (3), in fact, is indifferent between several candidate rankings, since there exist those pairs or groups of monkeys with no dominance paths between them. As a linear ranking is inappropriate in a situation such as this, we focus solely on inferring the SBT hierarchical structure of monkeys.

We turn to analyzing the hierarchical SBT power structure of the subjects within the network. In the

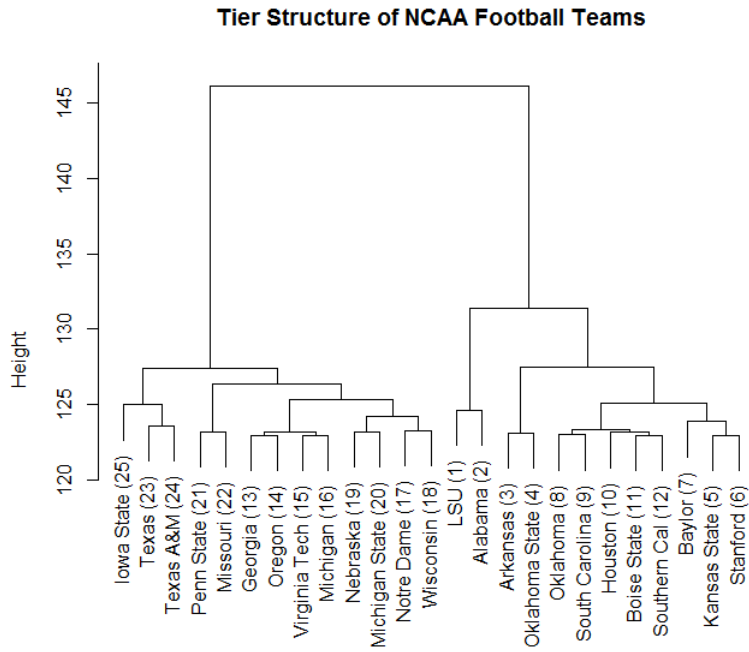


Figure 5: The hierarchical clustering tree of the top 25 teams showing the division of teams into tiers, using the matrix on left in Figure (4) as a distance matrix.

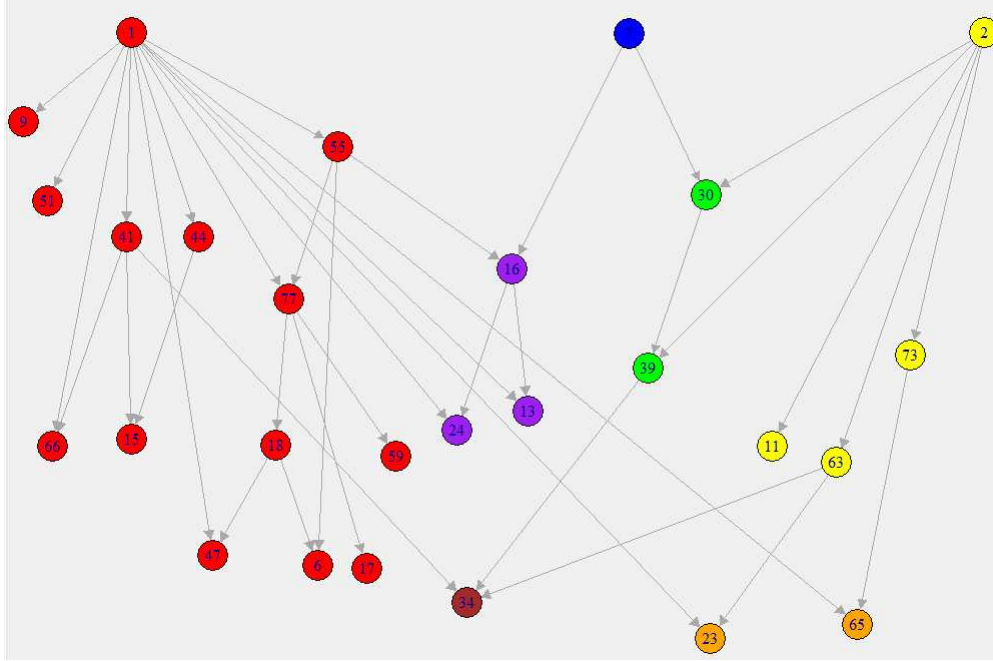


Figure 6: The hierarchical power structure of all monkeys dominating at least one other monkey with SBT actions. We focus on monkeys 1, 2, and 7; these monkeys were non-related males introduced to the cage after all other monkeys. Dominance sub-graphs are shown in red, yellow, and blue for monkeys 1, 2, and 7, respectively. Nodes contained in multiple sub-graphs are displayed in secondary colors: orange nodes are dominated by monkeys 1 and 2, purple nodes are dominated by monkeys 1 and 7, and green nodes are dominated by monkeys 2 and 7. Brown nodes are dominated by all three non-related males.

network of monkey SBT actions, we notice that we may rank certain subsets of monkeys in a linear fashion, but in general, a full linear ranking is inappropriate. Dominance paths may be used to identify which subsets of subjects may be linearly ranked. Combining all dominance path information gives us a hierarchical power structure, which can be graphically displayed as in Figures (6) and (7).

The absence of cycles makes the hierarchical SBT power structure extremely clear. In Figure (6), we can clearly see that there is a deterministic network structure defined by silent SBT actions. In particular, we notice that subjects 1, 2, and 7 all seem to be highly respected monkeys, each with some level of SBT dominance over a sizeable number of other monkeys. We also observe that each of these non-related male monkeys commands respect over a different subset of the others, which may be indicative of clique-like behavior. However, these subsets overlap substantially.

We further notice the absence of cycles in the hierarchy in Figure (6). This characteristic of the directed graph indicates that there is very strong dominance information contained in each SBT action. In fact, SBT actions seem to be perfectly transitive. That is, dominance paths of any order in this network carry the same amount of dominance information as directly observed SBT actions.

In Figure (7), we see the full hierarchical structure of all monkeys dominated at some level by monkey 1, 2, or 7 in terms of SBT actions. Again, the absence of cycles allows the hierarchical SBT power structure of the monkey society to be easily constructed. We can more clearly see the range of influence that monkey 1 has in Figure (7). Shown in Figure (8) is the complete monkey SBT hierarchy organized by matriline. We observe not only a hierarchy of individual monkeys, but also some hierarchical structure of families or

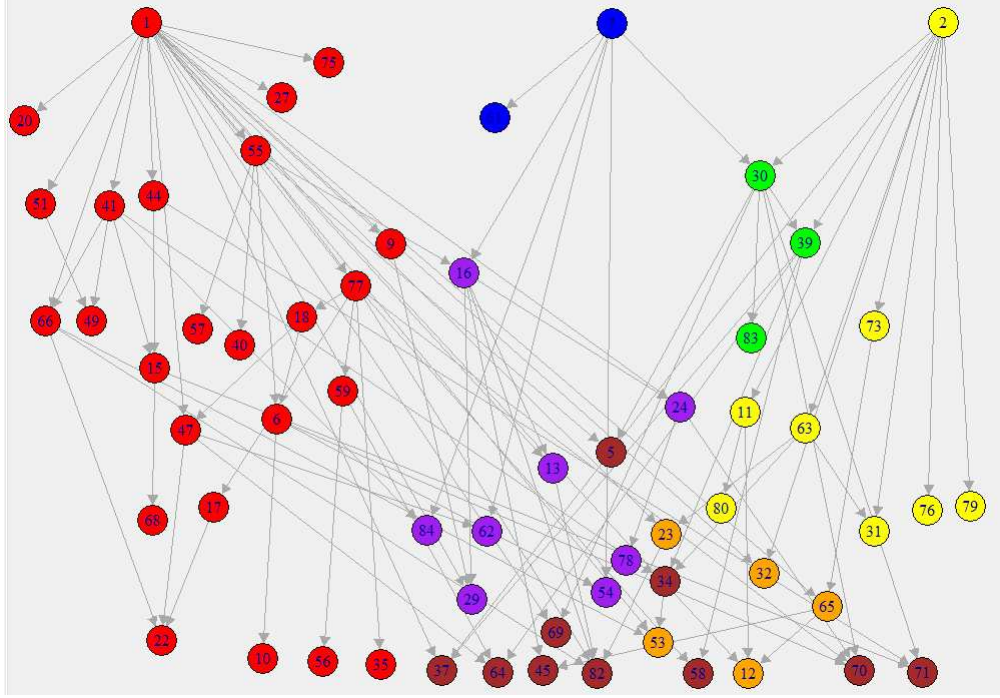


Figure 7: The hierarchical power structure of all monkeys dominated by monkey 1, 2, or 7.

groups of monkeys. Given the perfect transitivity and the deterministic hierarchical structure of the SBT network on an individual level as well as a familial level, it appears that silent SBT actions are important to consider when determining the monkeys' overall social standing. Because there is simply so little variability in the SBT network, it is much less likely that SBT actions are determined by some other measurable attribute of the monkey society. This is consistent with some primatologists' interpretation of the SBT as a formal subordination signal whose communication permits variability in the actual aggressive interactions of the dyad (de Waal, 1986; Flack and de Waal, 2003). The rigid SBT network structure provides us with some evidence that SBT actions are a primary structure in the monkeys' social group, but more research is necessary to conclude a causal relationship between SBT actions and other actions observed among the monkeys.

Other applications of social hierarchy inference

We may use the hierarchical structure inferred from a network to expose other characteristics of the society by involving other networks on the same subjects. While no other directed network is available for the NCAA football season, actions of aggression, alliance, and grooming were recorded on the monkeys discussed in the previous section. We will be using the SBT network's inferred structure in comparison with networks involving aggression, alliance, and grooming actions to examine possible relationships between the directed actions. Of particular interest is the connection between the SBT network with the three other networks: how are those different types of actions dependent on the monkeys' status in the SBT hierarchy, and are there certain relationships between monkeys that make a particular pair more or less likely to act in a certain manner toward each other?

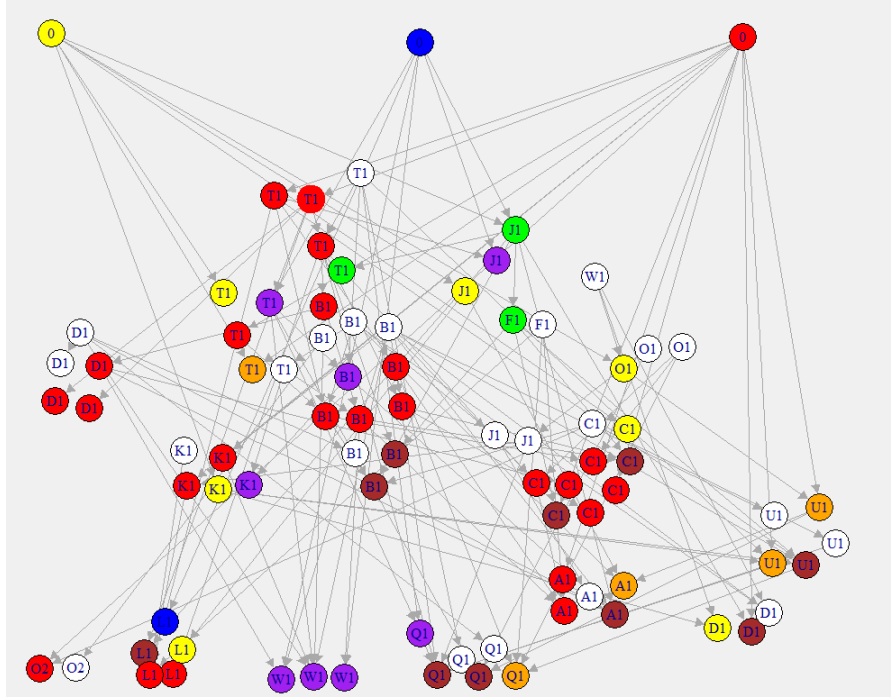


Figure 8: A graph of all monkey SBT actions, organized by matriline, so that all dominance actions are direct downwards in the graph. White-colored nodes are not dominated by any of the three non-related male monkeys.

Action	$i > j$	$i < j$	$i > j$ or $i < j$	Cannot determine relationship.
Aggression	1.726 (0.115)	0.193 (0.033)	1.919 (0.123)	1.338 (0.034)
Alliance	0.264 (0.035)	0.078 (0.018)	0.341 (0.043)	0.215 (0.012)
Grooming	0.166 (0.034)	0.291 (0.049)	0.456 (0.070)	0.429 (0.025)

Table 1: Average number of aggression, alliance, and grooming actions between two monkeys, conditional on their relative SBT standing. The direction of aggression and alliance actions seem to be more consistent with the direction of SBT actions, while grooming actions tend to occur in the opposite direction of SBT.

To address the first question, we refer back to the hierarchy displayed in Figure (7), implied by the SBT network. Table (1) shows the average number of each type of action, depending on how the monkeys are related in the SBT hierarchy. We calculate that, if monkey i is above monkey j in the SBT hierarchy, there is an average of 1.726 (± 0.115) aggression actions initiated by i upon j , and an average of 0.193 (± 0.033) aggression actions initiated by j upon i . This stark difference in prevalence of aggression actions leads us to infer that the monkeys are aware of their relative standing with respect to SBT, and observe the physical or social power of those monkeys higher up in the SBT hierarchy by not acting aggressively towards them. We notice a similar pattern in alliance actions, in which monkeys with a higher SBT standing are more likely to initiate an alliance with a monkey lower in the SBT hierarchy. However, we notice the opposite pattern with grooming actions: monkeys of lower SBT standing are more likely to groom those of higher SBT standing, perhaps as an act of submission.

To further examine the relationships between the networks, we can use the conductance algorithm described earlier to calculate “dominance probability” matrices for the aggression, alliance, and grooming

networks. The entries of these matrices represent the probability that, if an action has occurred between a specific pair of monkeys, it will have occurred in a particular direction. Comparing these matrices with the SBT dominance probability matrix will give us some insight into the relationship between the direction of the four actions that are recorded on the society of monkeys. When monkey i is higher in the SBT hierarchy than monkey j , we see that monkey i has a 90.54% chance of being more likely to initiate an act of aggression upon monkey j , rather than j upon i . Similarly, monkey i has an 83.96% chance of being more likely to initiate an act of alliance, and a 27.70% chance of being more likely to initiate an act of grooming, rather than in the other direction. This tells us that monkeys more highly regarded in an SBT sense are more likely to initiate aggression and alliance actions, but are less likely to groom others. That is, the direction of SBT, aggression, and alliance actions are typically consistent. However, SBT and grooming actions tend to occur in opposite directions, meaning that monkeys with lower SBT standing tend to groom those of higher SBT standing.

We also want to determine if the presence of a direct or indirect connection between a pair of monkeys in the SBT network is indicative of the presence of an aggression, alliance, or grooming action between those monkeys. If one monkey clearly dominates the other with regard to SBT (that is, there is positive SBT conductance in either direction), we observe 1.919 (± 0.123) aggression actions between the two, on average. If the relationship between monkey i and monkey j cannot be determined by SBT (that is, if there is no SBT conductance between i and j in either direction), there is only an average of 1.338 (± 0.034) aggression actions between that pair of monkeys.

Thus, by taking into account the hierarchy implied by the monkeys' SBT actions, we can better understand other networks, and therefore, the society of monkeys as a whole. Knowing how two monkeys relate socially in terms of SBT can help determine how likely they are to be aggressive toward each other, form an alliance, or groom one another.

Resampling on directed networks

We further use these methods to illustrate the futility in resampling on directed networks. The concept of bootstrapping or resampling methods revolve around making changes on a small scale to an observed data set, while holding the overall characteristics of the data set constant. For example, the parametric bootstrap estimates the distribution's parameters from the observed data, then resamples from that estimated distribution. The bootstrapped observations are slightly different, but the distribution and its parameters remain more or less unchanged.

In most cases, directed networks do not allow for such resampling actions to occur because any small change to an edge or node affects the network's global structure. In a conflict network such as the two we have worked with, directed edges carry dominance information. Moving, adding, deleting, or switching the direction of any edge can drastically alter the flow of such information. To illustrate these effects, we examine the changes in rankings of NCAA football teams when the directions of some edges are swapped.

We have identified 45 of the 677 football games whose outcomes we have swapped to create our resampled network. This new network has the same degree sequence (i.e., the in- and out-degrees of every node remain unchanged) and its edges still occur between the same pairs of teams. Upon recalculating conductivity between the 120 NCAA football teams, we notice some major differences in the competitive structure.

Shown in Figure (9) are the changes to the estimated dominance probability matrix when the resampled conflict matrix is used instead of the original conflict matrix. As can be seen, changes to the original

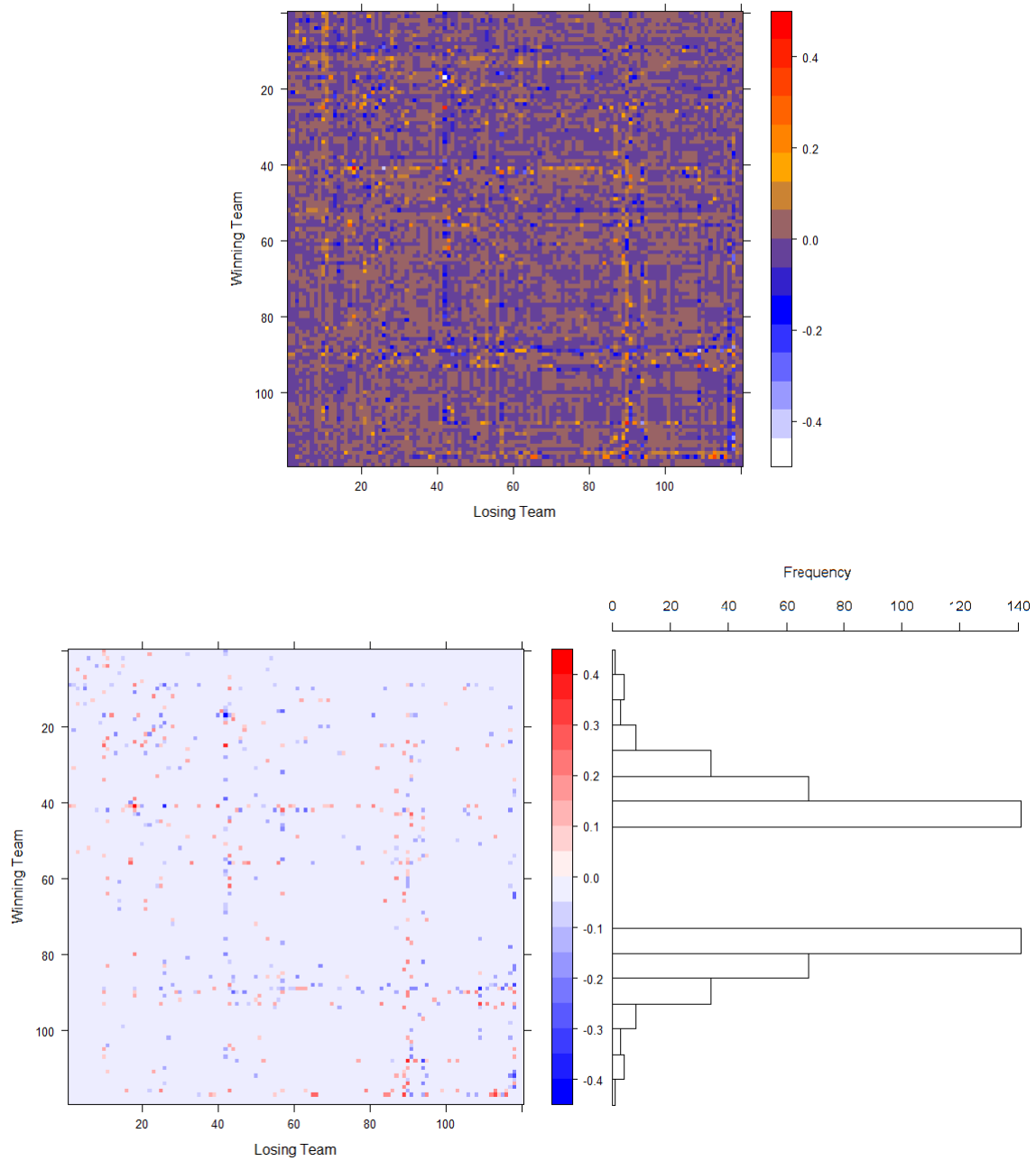


Figure 9: Differences in estimated dominance probabilities between the resampled conflict matrix and the original conflict matrix (top), as computed by Method 1. Cells corresponding to games whose results have been reversed have been set to zero, to further emphasize indirect effects of attempting to resample on the directed conflict network. Cells containing dominance probabilities that have changed by more than 0.1 are displayed on bottom. The histogram on right shows the distribution of the differences in estimated dominance probabilities incurred by resampling.

conflict matrix can have large effects to the rest of the network, due to diversions in the flow of dominance information. We notice that as a result of reversing the 45 game outcomes, 304 pairs of estimated dominance probabilities change by more than 0.1.

Such large differences in the estimated dominance probability matrix certainly affect ranking and hierarchical power structure inference. Table (2) compares the top 25 teams from the original conflict matrix and the resampled conflict matrix. Although the teams’ win-loss records have not changed, there are some major differences in the rankings. Houston’s ranking drops from 10 to 25, while Oregon and Stanford each dropped two spots from 3-4 ranks to 5-6 ranks. Meanwhile, Wisconsin’s rank rises from 22 to 11 and TCU goes from being unranked to 16th. Such differences in final rankings would likely have major implications when deciding which teams earn the right to compete in more prestigious bowl games, which would have large impacts on the schools involved due to financial incentives, publicity for the school and athletic program, and recruitment power.

Rank	Original Ranking	Bootstrapped Ranking
1	LSU	LSU (± 0)
2	Alabama	Alabama (± 0)
3	Oregon	Arkansas (+2)
4	Stanford	Oklahoma State (+2)
5	Arkansas	Oregon (-2)
6	Oklahoma State	Stanford (-2)
7	Kansas State	Southern Cal (+6)
8	Baylor	Kansas State (-1)
9	Oklahoma	Baylor (-1)
10	Houston	Oklahoma (-1)
11	Boise State	Wisconsin (+11)
12	South Carolina	South Carolina (± 0)
13	Southern Cal	Texas A&M (+7)
14	Missouri	Texas (+1)
15	Texas	Boise State (-4)
16	Georgia	TCU (+10)
17	Clemson	Georgia (-1)
18	Virginia Tech	Virginia Tech (± 0)
19	Auburn	Clemson (-2)
20	Texas A&M	Michigan State (+11)
21	Iowa State	Auburn (-2)
22	Wisconsin	Michigan (+1)
23	Michigan	Mississippi State (+6)
24	Texas Tech	Nebraska (+3)
25	Tulsa	Houston (-15)

Table 2: Top 25 NCAA football teams according to the original conflict matrix and the bootstrapped conflict matrix. Changes in rankings are shown in parentheses.

Discussion

As we have seen in the two examples, the use of direct and indirect dominance information in conjunction with each other can help refine estimates of dominance probabilities of one subject over another. More importantly, though, we can use both types of dominance information to infer the power structure of a

network, which can vary widely in its geometry. The approach for inferring the power structure with the proposed conductance algorithm depends largely on the network’s transitivity. Transitivity is key for not only determining the strength of the indirect dominance information available, but also to what degree it is possible, if at all, to use bootstrapping methods on a directed network.

When transitivity is approximately equal to 1, as we have seen with the monkey SBT data, the directed graph itself is the power structure. Even the slightest change to the network dramatically alters the power structure, so there is no room for randomization, or in particular, bootstrapping. In such cases, we may essentially refrain from identifying dominance paths, since the same information is contained in the observed conflicts. Instead, we can invest computational power in graphically rearranging the network, as in Figures (6) and (7).

If transitivity is moderate, though, as with the NCAA football data, dominance paths carry less dominance information, and thus it is generally less important to preserve each individual dominance path. The power structure is still discernable, but since the transitivity is less than perfect, is no longer perfectly represented by the observed network. Thus, computational power needs to be invested in identifying dominance paths and estimating dominance probability matrices. Still, the room for bootstrapping is small. We observed that even resampled networks that preserved the degree sequence in the NCAA football network did not preserve the power structure, as rankings were volatile.

It is an interesting case when transitivity is less than or equal to 0.5, and we aim to further explore this case in the future. In such a situation, order-1 dominance paths contain no information about dominance (when transitivity is equal to 0.5) or could even imply submission (when transitivity is less than 0.5). In competitive societies, we clearly expect transitivity to be greater than 0.5, but evidence otherwise would imply an intriguing competitive dynamic.

References

- Beisner, B. A., Jackson, M. E., Cameron, A. N., McCowan, B., 2011. Detecting instability in animal social networks: genetic fragmentation associated with social instability in rhesus macaques. *PLoS One* **6**(1): e16365. doi:10.1371/journal.pone.0016365.
- Bradley, R. A., Terry, M. E., 1952. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika* **39** (3/4), 324-345.
- de Waal, F. B. M., 1986. The integration of dominance and social bonding in primates. *The Quarterly Review of Biology* **61** (4), 459-479.
- de Waal, F. B. M., Luttrell, L. M., 1985. The formal hierarchy of rhesus monkeys: An investigation of the bared-teeth display. *American Journal of Primatology* **9**, 73-85.
- Flack, J. C., de Waal, F. B. M., 2003. Dominance style, social power, and conflict management in macaque societies: A conceptual framework (SFI working paper: 2003-06-040). URL <http://www.santafe.edu/media/workingpapers/03-06-040.pdf>.
- Flack, J. C., de Waal, F. B. M., 2007. Context modulates signal meaning in primate communication. *Proceedings of the National Academy of Sciences* **104**, 1581-1586. doi:10.1073/pnas.0603565104.
- Fushing, H., McAssey, M. P., McCowan, B., 2011. Computing a ranking network with confidence bounds from a graph-based Beta random field. *Proceedings of the Royal Society A* doi: 10.1098/rspa.2011.0268.
- Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., 1983. Optimization by simulated annealing. *Science* **220** (4598), 671-680.

McCowan, B., Beisner, B. A., Capitanio, J. P., Jackson, M. E., Cameron, A. N., Seil, S., Atwill, E. R., Fushing, H., 2011. Network stability is a balancing act of personality, power, and conflict dynamics in rhesus macaque societies. *PLoS One* **6**(8): e22350. doi:10.1371/journal.pone.0022350.

Nikolić, D., 2007. Non-parametric detection of temporal order across pairwise measurements of time delays. *Journal of Computational Neuroscience* **22** (1), 5-19.

Preuschoft, S., 1995. 'Laughter' and 'smiling' in macaques: an evolutionary perspective. PhD thesis, Rijkuniversiteit Utrecht, Utrecht.

Simon, H. A., 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* **106**, 467-482.

Appendix A: Conflict matrix imputation for Method 1

We identify all dominance paths up to a reasonable order M , and use them to fill dominance path matrices D_1, \dots, D_M . To emulate the construction of the ensemble matrix E from the percolation algorithm, we instead construct the non-stochastic ensemble matrix E^* , whose entries e_{ij}^* are calculated by

$$e_{ij}^* = c_{ij} + \sum_{m=1}^M \sum_{k=1}^{K_m} \prod_{f=0}^m \left(\frac{\alpha c_{\sigma_k^m(f)\sigma_k^m(f+1)} + \beta}{\bar{w}(\alpha(c_{\sigma_k^m(f)\sigma_k^m(f+1)} + c_{\sigma_k^m(f+1)\sigma_k^m(f)}) + 2\beta)} \right), \quad (4)$$

where $\bar{w} = \frac{1}{N} \sum_i w_i$ is the average number of observed dominance actions per subject. We include \bar{w} in the ensemble matrix calculation as a correction for the number of imputed dominance actions, since the number of dominance paths of any order increases exponentially with \bar{w} . The inclusion of \bar{w} in the above calculation also helps to ensure that the estimation of the dominance probability matrix is consistent.

The calculation in Equation (4) simplifies to

$$e_{ij}^* = c_{ij} + \sum_{m=1}^M d_{ij}^m \left(\frac{\alpha + \beta}{\bar{w}(\alpha + 2\beta)} \right)^{m-1}, \quad (5)$$

under the assumption that $c_{ij} + c_{ji} \leq 1$.

Consider running the percolation algorithm one time through, but suppose that we have already identified a particular dominance path, σ_k^m , that we will attempt to complete without terminating the algorithm. Let us simplify the percolation further by fixing the probabilities of success of the Bernoulli random variables X_f as the means of their respective Beta posterior distributions

$$\text{Beta}(\alpha c_{\sigma_k^m(f)\sigma_k^m(f+1)} + \beta, \alpha c_{\sigma_k^m(f+1)\sigma_k^m(f)} + \beta)$$

so that

$$P(X_f = 1) = \left(\frac{\alpha c_{\sigma_k^m(f)\sigma_k^m(f+1)} + \beta}{\bar{w}(\alpha(c_{\sigma_k^m(f)\sigma_k^m(f+1)} + c_{\sigma_k^m(f+1)\sigma_k^m(f)}) + 2\beta)} \right).$$

The probability that we successfully complete the percolation along the dominance path σ_k^m is therefore

$$\prod_{f=0}^m P(X_f = 1) = \prod_{f=0}^m \left(\frac{\alpha c_{\sigma_k^m(f)\sigma_k^m(f+1)} + \beta}{\bar{w}(\alpha(c_{\sigma_k^m(f)\sigma_k^m(f+1)} + c_{\sigma_k^m(f+1)\sigma_k^m(f)}) + 2\beta)} \right)$$

Thus, we simply weight each dominance path by its long-run probability of being successfully completed in

a particular round of percolation.

Appendix B: Alternative methods for imputing missing dominance data and ranking

A regression approach

The second proposed method takes a regression approach to imputing dominance information. Intuitively, dominance paths of higher orders contain less dominance information than paths of lower order, due to transitivity of dominance being subject to random error at every step in the path. Empirically, we observe the relationship

$$D_M \approx a D_{M-1} \quad (6)$$

for some constant $a > 1$ and sufficiently high order M . This helps to confirm our belief that we gain less and less dominance information as the order of the dominance paths grows. Therefore, dominance paths of lower orders should have more “predictive power” of true dominance than dominance paths of higher orders.

In particular, we can use dominance paths to help predict the presence of an observed dominance action using linear (or logistic) regression. Let $T = \sum_{i,j} c_{ij}$ be the total number of conflicts observed between subjects in the network. For simplicity of notation, we re-enumerate the T conflicts by c_1, \dots, c_T . Let the winner of conflict t be a_t and the loser be b_t . Then, we generate a sequence of T fair Bernoulli random variables, y_1, \dots, y_T . If $y_t = 1$, we set $x_{tm} = \frac{d_{a_t}^m}{d_{a_t}^m + d_{b_t}^m}$. If $y_t = 0$, we set $x_{tm} = \frac{d_{b_t}^m}{d_{a_t}^m + d_{b_t}^m}$ (if the denominator is zero, we set $x_{tm} = 0.5$). We then compose vectors of regressors $x_m = (x_{1m}, \dots, x_{Tm})'$ for $m = 1, \dots, M$, and a vector of observed wins or losses $y = (y_1, \dots, y_T)'$.

However, we note that because of the relationship in (6), issues of multicollinearity will be apparent for higher orders of m . Therefore, it is necessary to orthonormalize the regressors x_1, \dots, x_M using a procedure such as Gram-Schmidt. We refer to the orthonormalized regressors as u_1, \dots, u_M and proceed with the linear regression.

As previously mentioned, we may use logistic regression rather than linear regression. The main drawback to this approach is that estimated coefficients do not remain fixed when orthogonal regressors are added or removed from the model. Therefore, we still have issues with multicollinearity at high orders of dominance paths. However, logistic regression may be an attractive approach, as it guarantees that fitted values lie between 0 and 1, which are easy to interpret as the estimated probability of the presence of dominance action.

Regression allows us to easily compute fitted values $\hat{x}_1, \dots, \hat{x}_T$ for the T sets of observations included in the regression. Because of the orthonormalization of the regressors, though, it is no longer an easy task to compute fitted values for the other sets of observations. However, we can still estimate them. Let $x_t^* = (x_{t1}, \dots, x_{tM})'$ be observation t in the regression, and let $z_{ij} = \left(\frac{d_i^1}{d_i^1 + d_j^1}, \dots, \frac{d_i^M}{d_i^M + d_j^M} \right)'$ be the observation for which we would like to estimate a fitted value. Calculate the correlation between z_{ij} and x_t^* for $t = 1, \dots, T$. For the value t^* of t that maximizes this correlation, assign the fitted value \hat{x}_{t^*} to z_{ij} . We can then construct a fitted value matrix $Z = [z_{ij}]$, and finally treat $Z + C$ as the observed conflict matrix. A ranking may be obtained by again using simulated annealing with the cost function in (3).

An attractive feature of a regression approach, especially with linear regression, is that the coefficient estimates give us information about the importance of dominance paths of various orders. If we observe a

quick decay in coefficient estimates as the dominance path order rises, then we can infer that dominance paths of higher order are likely not important when imputing dominance information. However, if the decay in coefficient estimates is slower, then those dominance paths of higher order may still contain a significant amount of information. If we do not observe any decay in coefficient estimates, then we may instead need to explore what may be causing higher-order dominance paths to be as or more important than lower-order paths. Such an observation implies a violation in transitivity of dominance information, which in itself may be interesting to study.

A conductance algorithm for linear ranking

We can simplify the use of conductance to provide a simple linear ranking of subjects. In particular, we use dominance paths to assign dominance ratings to the subjects. This is reminiscent of the Bradley-Terry model (Bradley and Terry, 1952) for ranking. Instead of maximizing a likelihood function, though, we consider minimizing a loss function based on a proposed set of dominance ratings, given the set of dominance paths up to order M . Due to the transitivity of dominance information, we expect that more dominance paths exist from subjects with high dominance potential to low dominance potential, than vice versa. Thus, our objective is to choose the set of dominance ratings such that the number of dominance paths from a low-rated subject to a high-rated subject is minimized. We denote the vector of dominance ratings by $\mathbf{r} = (r_i)_{i=1}^N$. For a particular dominance path σ_k^m , consider the loss function

$$L(\sigma_k^m, \mathbf{r}) = \left(\frac{\alpha + \beta}{\bar{w}(\alpha + 2\beta)} \right)^{m-1} e^{-(r_{\sigma_k^m(0)} - r_{\sigma_k^m(m+1)})}. \quad (7)$$

We recognize the first component of the loss function in (7) as the weight assigned to dominance paths of order m from (5). Dominance paths of lower order contain more dominance information than paths of higher order. Therefore, we place more importance on lower-order dominance paths via this weight. The latter portion of (7) takes into account the difference in ratings between the starting node and the ending node in the dominance path σ_k^m . If the starting node $\sigma_k^m(0)$ is rated higher than the ending node $\sigma_k^m(m+1)$, then the penalty is very slight. However, if $\sigma_k^m(0)$ is rated lower than $\sigma_k^m(m+1)$, the penalty grows exponentially with the difference in ratings.

We further note that the loss function in (7) is always positive, but is smallest when the difference in ratings $r_{\sigma_k^m(0)} - r_{\sigma_k^m(m+1)}$ is as large as possible. By penalizing every dominance path, even by the slightest amount, we further promote the high rating of those subjects with an exceptional number of dominance paths emanating from them.

Taking all dominance paths of order up to M into account, our overall loss function is given by

$$L(\mathbf{r}) = \sum_{m=1}^M \sum_{k=1}^{K_m} L(\sigma_k^m, \mathbf{r}) = \sum_{m=1}^M \sum_{k=1}^{K_m} \left(\frac{\alpha + \beta}{\bar{w}(\alpha + 2\beta)} \right)^{m-1} e^{-(r_{\sigma_k^m(0)} - r_{\sigma_k^m(m+1)})}, \quad (8)$$

which is minimized with the use of simulated annealing to obtain an optimal set of dominance ratings.

We acknowledge that, as with the Bradley-Terry model, this method provides information about ranking, but none about a hierarchical power structure. However, it improves on the Bradley-Terry model for ranking by incorporating indirect dominance information in the optimization. The choice of α and β further allow flexibility on a context-dependent basis.

This method also provides a “sanity test” as reassurance that it is appropriate to directly consider

dominance paths when interpolating dominance information. As shown in the Results section, these three methods dependent on dominance paths provide reasonable rankings of NCAA football teams, but allow for further inference.

Appendix C: Top 25 NCAA football teams as determined by our algorithms

Rank	BCS Ranking	Conductance Ranking	Regression Ranking	Dominance Rating Ranking
1	LSU (12-0)	LSU (12-0)	LSU (12-0)	LSU (12-0)
2	Alabama (10-1)	Alabama (10-1)	Alabama (10-1)	Alabama (10-1)
3	Oklahoma State (11-1)	Arkansas (9-2)	Stanford (11-1)	Oklahoma State (11-1)
4	Stanford (11-1)	Oklahoma State (11-1)	Arkansas (9-2)	Stanford (11-1)
5	Oregon (10-2)	Kansas State (9-2)	Kansas State (9-2)	Arkansas (9-2)
6	Arkansas (9-2)	Stanford (11-1)	Southern Cal (10-2)	Kansas State (9-2)
7	Boise State (11-1)	Baylor (8-3)	Wisconsin (10-2)	Boise State (11-1)
8	Kansas State (9-2)	Oklahoma (9-3)	Oklahoma State (11-1)	South Carolina (9-2)
9	South Carolina (9-2)	South Carolina (9-2)	Baylor (8-3)	Oregon (10-2)
10	Wisconsin (10-2)	Houston (11-1)	Oklahoma (9-3)	Oklahoma (9-3)
11	Virginia Tech (10-2)	Boise State (11-1)	Virginia Tech (10-2)	Baylor (8-3)
12	Baylor (8-3)	Southern Cal (10-2)	Boise State (11-1)	Georgia (9-3)
13	Michigan (9-2)	Georgia (9-3)	Georgia (9-3)	Virginia Tech (10-2)
14	Oklahoma (9-3)	Oregon (10-2)	Auburn (6-5)	Southern Cal (10-2)
15	Clemson (9-3)	Virginia Tech (10-2)	Missouri (6-5)	Wisconsin (10-2)
16	Georgia (9-3)	Michigan (9-2)	Oregon (10-2)	Houston (11-1)
17	Michigan State (9-3)	Notre Dame (8-4)	Texas A&M (6-6)	Michigan (9-2)
18	TCU (9-2)	Wisconsin (10-2)	Michigan State (9-3)	Auburn (6-5)
19	Houston (11-1)	Nebraska (8-3)	Texas (7-5)	Penn State (8-3)
20	Nebraska (8-3)	Michigan State (9-3)	Texas Tech (4-7)	TCU (9-2)
21	Southern Miss (10-2)	Penn State (8-3)	Penn State (8-3)	Michigan State (9-3)
22	Penn State (8-3)	Missouri (6-5)	Michigan (9-2)	Texas (7-5)
23	West Virginia (8-3)	Texas (7-5)	Nebraska (8-3)	Texas A&M (6-6)
24	Texas (7-5)	Texas A&M (6-6)	Iowa State (5-6)	Tulsa (8-4)
25	Auburn (6-5)	Iowa State (5-6)	Tennessee (4-7)	Clemson (9-3)

Table 3: Top 25 NCAA football teams as ranked by the BCS and the three proposed methods.

Appendix D: R Code for identifying dominance paths

```
IDpaths = function(conf, i, len){
  ### IDpaths: identifies all unique dominance paths of order (len - 1) beginning at subject i
  ### conf: N-by-N conflict matrix whose (i,j)th element is the number of times i defeated j
  ### i: the subject at the beginning of each dominance path
```

```

    ### len: the length of the dominance paths to be identified (len = order + 1)

if(sum(conf[i,] > 0) == 0) return(NULL)
levels = list()
levels[[1]] = which(conf[i,] > 0)
for(j in 2:len){
  levels[[j]] = lapply(unlist(levels[[j-1]]), function(k) which(conf[k,] > 0))
}
ret = matrix(0, length(unlist(levels[[len]])), len+1)
ret[,1] = i
ret[,len+1] = unlist(levels[[len]])
if(len == 2){
  ret[,2] = rep(unlist(levels[[1]]), sapply(levels[[2]], length))
}
for(j in len:2){
  currLengths = sapply(levels[[j]], length)
  if(j < len){
    effLengths = numeric(length(currLengths))
    ctr = 1
    for(d in 1:length(effLengths)){
      if(currLengths[d] != 0){
        effLengths[d] = sum(prevLengths[ctr:(ctr + currLengths[d] - 1)])
      }
      else{
        effLengths[d] = 0
      }
      ctr = ctr + currLengths[d]
    }
  }
  else{
    effLengths = currLengths
  }
  ret[,j] = rep(unlist(levels[[j-1]]), effLengths)
  prevLengths = effLengths
}
isUnique = apply(ret, MARGIN = 1, function(b) {length(unique(b)) == len + 1})
ret[isUnique,]      ### return only those dominance paths containing no cycles
}

```