frontiers
in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 28 July 2016
doi: 10.3389/fams.2016.00009

Mimicking Directed Binary Networks for Exploring Systemic Sensitivity: Is NCAA FBS a Fragile Competition System?

Fushing Hsieh * *and Kevin Fujii*

Department of Statistics, University of California, Davis, Davis, CA, USA

OPEN ACCESS

Edited by:
Haiyan Cai,
University of Missouri-St. Louis, USA

Reviewed by:
Ashley Prater,
United States Air Force Research Laboratory, USA
Roza Aceska,
Ball State University, USA

Correspondence:
Fushing Hsieh
fhsieh@ucdavis.edu

Specialty section:
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

Received: *27 April 2016*
Accepted: *13 July 2016*
Published: *28 July 2016*

Citation:
Hsieh F and Fujii K (2016) Mimicking Directed Binary Networks for Exploring Systemic Sensitivity: Is NCAA FBS a Fragile Competition System? Front. Appl. Math. Stat. 2:9. doi: 10.3389/fams.2016.00009

Can a popular real-world competition system indeed be fragile? To address this question, we represent such a system by a directed binary network. Upon observed network data, typically in a form of win-and-loss matrix, our computational developments begin with collectively extracting network's information flows. And then we compute and discover network's macrostate. This computable macrostate is further shown to contain deterministic structures embedded with randomness mechanisms. Such coupled deterministic and stochastic components becomes the basis for generating the microstate ensemble. Specifically a network mimicking algorithm is proposed to generate a microstate ensemble by subject to the statistical mechanics principle: All generated microscopic states have to conform to its macrostate of the target system. We demonstrate that such a microstate ensemble is an effective platform for exploring systemic sensitivity. Throughout our computational developments, we employ the NCAA Football Bowl Subdivision (FBS) as an illustrating example system. Upon this system, its macrostate is discovered by having a nonlinear global ranking hierarchy as its deterministic component, while its constrained randomness component is embraced within the nearly completely recovered conference schedule . Based on the computed microstate ensemble, we are able to conclude that the NCAA FBS is overall a fragile competition system because it retains highly heterogeneous degrees of sensitivity with its ranking hierarchy.

Keywords: beta random field, complex system, macrostate, network mimicking, system robustness

1. INTRODUCTION

Advances in Information Technology have equipped scientists with unprecedented capabilities to create new systems as well as to peek into old systems in human societies and in nature. Scientific endeavors of collecting and mining diverse kinds of high frequency, sequencing and network data for better understanding on these new and old systems of interest are nearly overwhelming across all branches of science. Though methodological techniques on computing and analyzing data derived from such systems have progressed in many fronts in past decades, still the speed of progress on learning from data apparently lags far behind the speed of generating data. One of the key reasons behind this lagging phenomenon can be seen from the perspective of order and chaos, as described in Crutchfield [1]. Any physical system of scientific interest is simultaneously characterized by its deterministic structures and its randomness. Critically these two systemic characteristics are coupled with each other in unknown and complex fashions. This coupling relation is hardly simple, much less linear which contributes to the complexity of real world systems.

Within a complex system, Crutchfield [1] asked questions: What is a pattern? How do we come to recognize and compute patterns never seen before? He further asserted that quantifying the notion of pattern and formalizing the process of pattern discovery go right to the heart of physical science. To formalize a suitable process of pattern discovery in a target system, some knowledge and insights regarding the architecture of systemic complexity, which have been advanced several decades ago, become even more valuable than ever. For instance, Simon [2] argued that hierarchical organizations likely merge in the process of evolution as systems converge toward their stable states, while Anderson [3] brought out that the twin difficulties of scale and complexity are keys to understanding large and complex systems. Up to now much more light has been shed on complex systems. A series of interesting related work can be found in the journal Chaos's Focus Issue on "Randomness, Structure, and Causality: Measure of Complexity from Theory to Applications, see Crutchfield and Machta [4] and papers therein.

In this paper we not only heel to the insights of hierarchical and multiscale structures when formalizing a process of pattern discovery upon a real world complex system, but also we take one step further to make use the discovered systemic patterns for exploring systems' sensitivity. That is, after computing and discovering deterministic structures in the form of pattern hierarchy, we attempt to explore which scales of the hierarchical constructs are susceptible to perturbations within discovered reign of randomness. This perspective of coupling deterministic and stochastic macroscopic components on research of directed network is still missing in literature of statistical mechanics on complex networks, see Albert and Barabási [5]. In contrast, recent works on universal resilience patterns in complex network are primarily studied through multiple dimensional differential equations, see Barzel and Barabási [6] and Gao et al. [7]. Our data-driven exploration endeavors here should add one more dimension of systemic understanding from the perspective of control theory, in which a system's complexity is thought of being characterized by its robustness as well [8].

Here our focus is particularly placed on any system that can be represented or approximated by a directed binary network. For the purpose of concrete and real exposition, but without loss of generality, we employ the competitive system of the NCAA Football Bowl Subdivision (FBS) to illustrate our computational developments. A glimpse at the characteristics of a directed binary network is given as follows.

It can be mathematically represented by a non-symmetric square binary matrix. The directedness refers to its relational "flows" on each linkage beginning from one node and ending on another node. A collection of such flows make up ample quantities of flow-paths, also called dominance paths, of various lengths. It likely also contains cyclic flow-paths as loops of different lengths. In the sense of dominance, a cyclic flow-path indicates a piece of conflicting dominance information. In the NCAA example, flow-path cycles are prevalent. It is worth noting that this case is not universal. For instance, the Rhesus Macaque monkeys' Silent Bared Teeth (SBT) behavioral networks, as reported in Fujii et al. [9], have no such flow-cycles. That is, there is no ambiguity in any dominance relationship in the monkey society through the SBT behavior. In contrast the NCAA Football competitive system is much more complex in this aspect, since it indeed allows highly heterogeneous degrees of uncertainties among all possible node-pairs. As would be seen in sections below, such conflicting information is the primary source of computational complexity and difficulty in directed binary networks. Though this fact has long been recognized by animal scientists, who studied ranking structures in animal societies (see [10–14]), so far no theories are available to constitute optimal methodologies for extracting global structural patterns from directed binary network data (see [15]).

For expositional simplicity within the competition system of NCAA football, a directed binary network is represented by a binary "win-and-loss" matrix. And the symmetrized "win-and-loss" matrix, the sum of the transposed "win-and-loss" matrix with itself, is a symmetric undirected schedule matrix.

We attempt the pattern discovery by following steps. We first construct a dominance probability matrix by extracting all pairwise dominance potentials from the collection of dominance paths contained within an observed binary "win-and-loss" matrix. We then take this dominance probability matrix as a thermodynamic system, on which we build a Hamiltonian. As such we quantify the deterministic structure by the minimum energy ground state, or macrostate. Such a computable macrostate reveals a power hierarchy among all involved nodes. Another pattern structure, called the Parisi adjacency matrix, is computationally discovered from the schedule matrix via an approach also based on statistical mechanics, see Fushing et al. [16]. We then argue that this dominance probability matrix together with the computed Parisi adjacency matrix constitute the discovered randomness as another intrinsic stochastic component embedded within the directed binary network under study. Based on such a randomness component, we are able to propose a network mimicking algorithm to generate microstates that are indeed conforming the macrostate. The resultant microstate ensemble would be used as an operational platform for exploring systemic sensitivity. This series of computational development is exclusively illustrated upon 2 years of NCAA college football data from the 2012 and 2013 seasons.

2. MATERIALS AND METHODS

2.1. Win-and-Loss Matrix

Let a directed binary network $\mathcal{G}(\mathcal{W}_0)$ be represented by a $m \times m$ binary win-and-loss matrix $\mathcal{W}_0 = [\mathcal{W}_0[i,j]]$ with $\mathcal{W}_0[i,j] = 1$ meaning team i having one "win" over team j, and $\mathcal{W}_0[i,j] = 0$ one "loss" to team j. The sum of $\mathcal{W}_0$ and its transpose, $\mathcal{A}_0 = \mathcal{W}_0 + \mathcal{W}_0^T$, is termed the schedule matrix. $\mathcal{A}_0$ is binary and symmetric.

The 2012 and 2013 NCAA college football schedules are shown in **Figures 1A,B**, respectively. There are 124(= m) teams in 2012, and 125 in 2013, competing in the two NCAA football examples. The majority of teams are divided into 10 conferences,

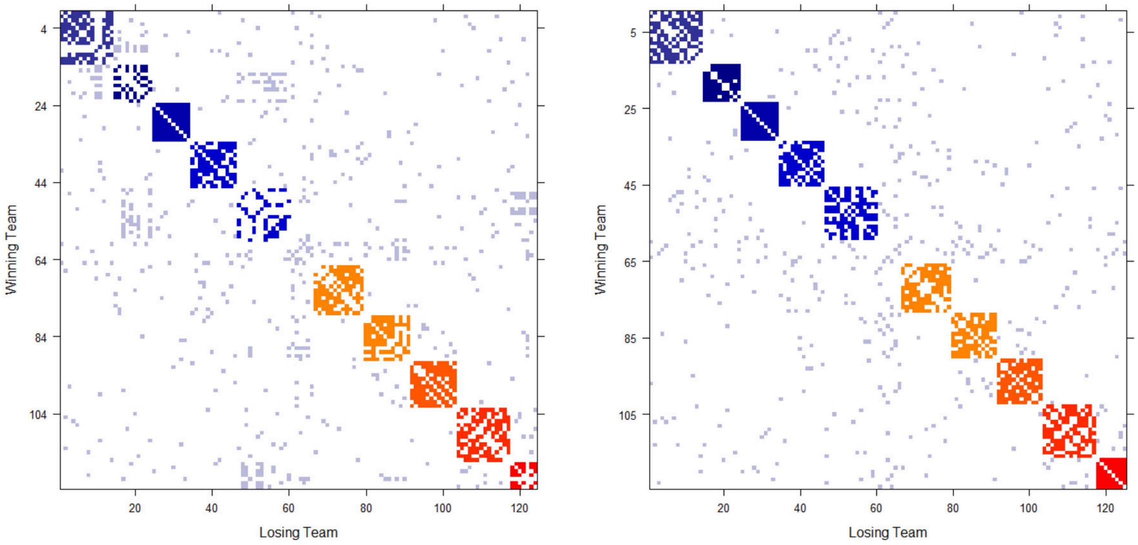

FIGURE 1 | Schedule matrix $\mathcal{A}_0$ of NCAA football data for the 2012 season (Left panel) and 2013 season(Right panel). The schedule matrix is composed of colored diagonal blocks marking the intra-conference games for the real conferences, and off-diagonal inter-conference games are marked in gray. The six independent teams account for the gap in the middle rows and columns. Conferences, from blue to red, are ACC, American, Big 12, Big Ten, Conference-USA, Independent schools, Mid-American, Mountain West, Pac 12, SEC, and Sun Belt.

such as Pac-12, Big-10, and SEC etc. with 6 independent teams. The ten block structures in **Figure 1** show more intra-conference than inter-conference games. Each team plays 12 or 13 games without multiplicity against any opponent team. That is, if two teams meet, we have $\mathcal{W}_0[i,j] + \mathcal{W}_0[j,i] = \mathcal{A}_0[i,j] = \mathcal{A}_0[j,i] = 1$, so that they meet only once in the regular season. As a result, these two win-and-loss matrices are not only authentically binary, but also are rather sparse since there are in total fewer than 750 games played among more than 7500 possible team-pairs. Hence nine out of ten entries in both matrices are missing. In sharp contrast a dominance probability matrix computed below solely based on the two $\mathcal{W}_0$ has nearly no missing entries.

2.2. Initial Beta Random Field

To begin, it is natural to ask: How much a win is worth in the sense of dominance? Formally, given the fact that $\mathcal{W}_0[i,j] = 1$, what is the conditional probability that team i would win against team j in their next hypothetical encounter? Such a probability surely depends on the nature or uncertainty in outcome of the "game," such as the NCAA football competitive landscape, as well as on their discrepancy of their winning potentials, which is intuitively related to the difference of their "true" ranking statuses. To mathematically formalize these thoughts, we propose to model such a conditional probability through a Beta random variable $Be(w[i,j]\alpha_{i,j} + 1, w[j,i]\alpha_{i,j} + 1)$. This proposal relies on the fact that $Be(w[i,j] + 1, w[j,i] + 1)$ is a Beta posterior distribution under a Bernoulli trial with Uniform prior $U[0,1] = Be(1,1)$. And $\alpha_{i,j}$ would be computed in a data-driven fashion. The larger $\alpha_{i,j}$ value is, the distribution of $Be(w[i,j]\alpha_{i,j} + 1, w[j,i]\alpha_{i,j} + 1)$ is more concentrated toward 1.

To compute $\alpha_{i,j}$ in a data-driven fashion, we consider the following decomposition $\alpha_{i,j} = \alpha_0 + \Delta\alpha_{i,j}$, where α_0 accounts for the nature of NCAA football games, and $\Delta\alpha_{i,j}$ for the discrepancy of winning potentials between teams i and j. The α_0 is computed based on the overall triad-based transitivity, while $\Delta\alpha_{i,j}$ is based on their tree distance found in a classification hierarchy of all involving teams.

The triad-based transitivity on $\mathcal{W}_0$ is computed as follows. Here an order-0, or a direct, dominance of (i,j) pair is referring to the case $w[i,j] + w[j,i] = 1$. If $w[i,j] = 1$, i is direct dominating j. An order-1 dominance path of (i,j) pair refers to the existence of one intermediate team-node k such that, either $w[i,k] = 1, w[k,j] = 1$ for the order-1 indirect dominance of i over j, or $w[j,k] = 1, w[k,i] = 1$ for j over i. Further a triad $\{i,j,k\}$ is termed being coherent in dominance if i dominates j, i dominates k and k dominates j. For a triad of nodes having three directed edges, there are only two possibilities: being coherent or incoherent between the direct and the order-1 indirect dominance direction. Therefore as one of global features of $\mathcal{W}_0$, the triad-based transitivity is defined as the proportion of coherent dominance, excluding all triads with fewer than three edges. The empirical proportion is denoted as $\hat{T}(\mathcal{W}_0)$.

Then we heuristically equate the transitivity estimate $\hat{T}(\mathcal{W}_0)$ to the dominance probability computed through an order-1 dominance path:

$$\hat{T}(\mathcal{W}_0) = \left(\frac{\alpha_0 + 1}{\alpha_0 + 2}\right)^2$$

where $\frac{\alpha_0+1}{\alpha_0+2}$ is the mean value of $Be(\alpha_0 + 1, 1)$.

Though this equation is rather simplistic, it works reasonably well when $\hat{T}(\mathcal{W}_0) \geq 0.5$. For instance, if the empirical $\hat{T}(\mathcal{W}_0)$ is calculated as 0.72, then the data-driven $\alpha_0 = 5.6$. Then the random variable $Be(5.6 + 1, 1)$ has its mean around 0.9 and variance less than 0.015. This context-dependent choice of Beta distribution is very different from $Be(2, 1)$ with its mean 2/3 and variance $0.056 (= 1/18)$ as used in Fushing et al. [17]. As the extreme case, if $T(\mathcal{W}_0) = 1.0$, the perfect transitivity, then $\alpha_0 = \infty$. It means that the dominance probability is always 1.

We place one Beta random variable with distribution $Be(\alpha_0 + 1, 1)$ onto every (i, j)-entry of the $m \times m$ lattice where $w[i, j] = 1$, otherwise the entry is left blank. It is emphasized that the blanks, not $Be(1, 1)$, are placed for entries without observed direct dominance, that is, for all (i, j) pairs with $w[i, j] + w[j, i] = 0$. The $m \times m$ matrix array of Beta random variables and blanks is called Beta Random Field in Fushing et al. [17].

This random field allows us to simulate random dominance paths mimicking the observed ones. This capability is critical for carrying out the "ensemble" idea for extracting the initial version of dominance probability matrix, from which the initial version of power structural hierarchy is derived, as will be discussed in the next subsection.

After the initial version of the power structural hierarchy is derived, tree distances for all pairs of nodes become available. We then estimate $\Delta\alpha_{i,j}$ for all (i, j) pairs having $w[i, j] + w[j, i] = 1$. Finally, refined versions of the dominance probability matrix and the power structural hierarchy are derived.

2.3. Initial Dominance Probability Matrix

The complexity of computing an initial dominance probability matrix, denoted as $P_D^{(0)}(\mathcal{W}_0)$ based on an observed win-and-loss matrix $\mathcal{W}_0$, involves one critical front: to identify all of the locations of "real structural" blank entries under a nonparametric framework. Here a "real structural" blank entry refers to a pair of nodes without dominance relationship, not because of lack of data. This task becomes particularly difficult when the number of all possible pairs $m(m - 1)/2$ is large and the number of missing entries, $w[i, j] + w[j, i] = 0$ is overwhelming in $\mathcal{W}_0$.

To resolve this critical computational issue, we devise an algorithm called the Trickling percolation process. Heuristically we make use of common percolation processes trickling within the Beta random field to explore and simulate as many dominance paths as possible. By performing a huge number of such percolation processes, we hope that collectively such percolating trajectories would be able to exhaustively visit every possible non-blank entries. Those entries completely missed by every single percolation processing should be reasonably taken as blank entries. At the same time the ensemble of simulated percolation trajectories is thought to achieve a uniform sampling of all possible dominance paths. Therefore any pair of nodes, which have dominance relationship, is sampled with pertinent frequencies in either dominance direction. The ratio of these two frequencies should properly reflect their dominance potential. The precise algorithmic description of the Trickling percolation process is given below.

2.3.1. Trickling Percolation Algorithm

TP-1: Consider a Markovian process to begin with a random initial node i_0 such that $\sum_{j=1}^{m} w_{i_0,j} > 0$ and make a zero $m \times m$ matrix E and index $k = 1$;

TP-2: Select a candidate neighbor i_k with equal probability from the set of nodes $\{i | w_{i_0,i} = 1, i = 1, .., m\}$(also excluding nodes already appeared in the Markovian process);

TP-3: Generate a random value p from $Be(\alpha_0+1, 1)$ and simulate a Bernoulli random variable $X_{i_k} \sim BN(p)$. If $X_{i_k} = 1$, then count it as a "win," otherwise a "loss." Record the win or loss as order-0 dominance for the node-pair into entries (i_0, i_k) and (i_k, i_0) in matrix E.;

TP-4: If $X_{i_k} = 1$, then add node i_k into the dominance path and let the Markovian process go to the TP-2 step with index $k = k + 1$. If it is loss, then a simulated dominance path ends, so is the Markovian process stops. Go to the next step;

TP-5: Transform all order-2, 3,.... of pairwise dominance along the Markovian simulated dominance path into order-0 pairwise dominance and record them into E.

Specifically the simulated dominance matrix E resulted from the Tricking percolation algorithm has 1's in the following collection of entries $\{(i_k, i_j)|k < j \leq K, k = 0, 1, ..., K\}$ when a realized simulated Markovian dominance path is the sequence $\{i_0, i_1, ...i_k, ...i_K, i_{K+1}\}$ with $X_{i_{K+1}} = 0$ and entries $E[i_K, i_{K+1}] = 0$ and $E[i_{K+1}, i_K] = 1$.

One simulation of the above Trickling percolation algorithm gives rise to one replication of simulated win-and-loss dominance matrix E. We construct a collection of a large number of replications of E's matrices, and denote the matrix summing over the win-and-loss dominance matrix ensemble as $\mathcal{E}(\mathcal{W}_0)$. We then take the following ensemble average via an entry-wise ratio operation: for the (i, j) entry of matrix $\overline{D}(\mathcal{W}_0)$,

$$\overline{D}(\mathcal{W}_0)[i, j] = \frac{\mathcal{E}(\mathcal{W}_\prime)[i, j]}{\mathcal{E}(\mathcal{W}_\prime)[i, j] + \mathcal{E}(\mathcal{W}_\prime)[j, i]}.$$

This special ensemble average $\overline{D}(\mathcal{W}_0)$ pragmatically provides reasonable pairwise dominance probability for all potential dominance relationships possibly embedded within $\mathcal{W}_0$. And very importantly, any blank in $\overline{D}(\mathcal{W}_0)$ reliably has no dominance relationships up to the extent of the percolating capability of Trickling percolation algorithm. Hence $\overline{D}(\mathcal{W}_0)$ is taken as a non-structural version of dominance probability matrix.

3. RESULTS

The computed dominance probability matrix $\overline{D}(\mathcal{W}_0)$ from the previous section would be the basis for extracting systemic structures: namely, its deterministic structure and its randomness. It is noted that the deterministic structure is likely a composite structure comprised of multiple overlapping trees. To bring out and present such a structural composition in one single hierarchy can be rather complicated. Here we represent two global aspects of the deterministic structure of dominance: one power-ordering axis and one power hierarchy. It is also noted that the system's randomness is highly associated with the deterministic structure and the environmental constraints pertaining to the system under study. Here we extract the pattern information contents from the schedule matrix $\mathcal{A}_0$. The systemic randomness is constructed by coupling a simulated schedule matrix, which conforms to the computed pattern constraints, with an estimate of the structured dominance probability matrix.

3.1. Systemic Structures I: Power Structural Hierarchy

The task of extracting NCAA college football's yearly deterministic structure involves two steps. First we derive an estimate of the $m \times m$ structural dominance probability matrix, denoted $P_D(\mathcal{W}_0)$, that reveals one global aspect of the dominance along a power ordering axis, which is not necessary in linear ordering. Second, based on the estimate of $P_D(\mathcal{W}_0)$, we derive a clustering tree that reveals another global aspect of dominance via a power hierarchy, in which nodes belonging a core cluster are viewed as not significantly different in dominance, while two nodes belonging to clusters on two different tree branches are considered significantly distinct in dominance. Indeed the larger tree distance between two nodes, the bigger discrepancy in dominance.

Intuitively an estimate of $P_D(\mathcal{W}_0)$ can be obtained by properly permuting rows and columns of $\overline{D}(\mathcal{W}_0)$ to reveal a power ordering axis. For extracting such a global structure out of $\overline{D}(\mathcal{W}_0)$, we consider the following Hamiltonian, or the cost function of the permuted matrix $U(\sigma)\overline{D}(\mathcal{W}_0)U^T(\sigma) = [d_{ij}(\sigma)]$ with permutation matrix $U(\sigma)$ pertaining to permutation $\sigma \in \mathcal{U}$:

$$\mathcal{H}(U(\sigma)\overline{D}(\mathcal{W}_0)U^T(\sigma)) = \sum_{i=2}^{m}\sum_{j=1}^{i-1}\max(0, -\log[2(1-d_{ij}(\sigma))])e^{\frac{(m+1-j)(i-j)}{m^2}}.$$

In the lower triangular matrix of $U(\sigma)\overline{D}(\mathcal{W}_0)U^T(\sigma)$, this Hamiltonian has a built-in function that not only penalizes entries larger than 0.5, but also penalizes such entries that are far away from the diagonal. Symmetrically in the upper triangular matrix, this corresponds to entries smaller than 0.5.

This Hamiltonian $\mathcal{H}(.)$ defines the potential field of $\{U(\sigma)\overline{D}(\mathcal{W}_0)U^T(\sigma)|\sigma \in \mathcal{U}\}$. Its ground state is taken as:

$$\mathcal{H}(U(\sigma^*)\overline{D}(\mathcal{W}_0)U^T(\sigma^*)) = \min_{\sigma \in \mathcal{U}} \mathcal{H}(U(\sigma)\overline{D}(\mathcal{W}_0)U^T(\sigma)).$$

It is known that the simulated annealing algorithm [18] is not effective to deal with computational complexity of searching for σ^* when m is not small. For moderate $m(\approx 100)$, this simulated annealing (SA) algorithm works effectively to reorder the rows and columns of $\overline{D}(\mathcal{W}_0)$ only when it starts on an initial state, which is already in the vicinity of the ground state. A reasonable initial state is simply the permutation obtained by sorting the row sums of $\overline{D}(\mathcal{W}_0)$. This choice of initial state seems to give rise to an efficient implementation of SA algorithm, particularly for the NCAA football data.

Let this SA algorithm provide one final permutation denoted by $\hat{\sigma}^*$, and one estimate of the $m \times m$ dominance probability matrix $P_D(\mathcal{W}_0)$ denoted by

$$\hat{P}_D(\mathcal{W}_0) = U(\hat{\sigma}^*)\overline{D}(\mathcal{W}_0)U^T(\hat{\sigma}^*).$$

The two computed dominance probability matrices for thee NCAA football data on years 2012 and 2013 are shown in the left panels of **Figures 2**, **3**. And we specifically term $\hat{\sigma}^*$ the "power ordering axis" of $\hat{P}_D(\mathcal{W}_0)$.

Next we address the task of extracting the power hierarchy embedded within $\hat{P}_D(\mathcal{W}_0)$. Here we build this power hierarchy via Ultrametric tree geometry based on a Hamiltonian distance matrix. Here, the new "distance" between two nodes is defined as the Hamiltonian incurred by switching the two nodes' positions in $\hat{\sigma}^*$ subtracting the minimum Hamiltonian $\mathcal{H}(U(\hat{\sigma}^*)\overline{D}(\mathcal{W}_0)U^T(\hat{\sigma}^*))$. Small values of such a distance indicate that these two nodes are indeed nearly equal in the power hierarchy, and should share a core cluster. By arranging the football teams on the new distance matrix according to $\hat{\sigma}^*$, as shown right panels of in **Figure 2** for Year 2012 and **Figure 3** for year 2013, the presence of a serial of small white blocks along the diagonal is clearly visualized. This block pattern information confirms that the permutation resulting from the simulated annealing algorithm is reasonable. Each white block indeed indicates a core cluster, in which all involved teams are considered equal in power or dominance.

Further core clusters, which only include nodes being very close to each others, should be allowed to merge into a conglomerate cluster upon a higher level of hierarchy. By discovering all involved levels, an Ultrametric tree geometry is built (see [19, 20] for the detailed algorithmic construction) called Data Cloud Geometry (DCG). It is essential to point out here that such a distance has a built-in global ingredient, which is implicitly critical in constructing an Ultrametric tree geometry. The two Ultrametric tree geometries pertaining to two years of NCAA college football are shown in **Figures 6**, **7**, respectively.

3.1.1. Refined Version of Power Structural Hierarchy

The estimation of the $m \times m$ dominance probability matrix $P_D(\mathcal{W}_0)$ via $\hat{P}_D(\mathcal{W}_0)$ derived above can be improved by considering the following version of local transitivity. We propose to compute local transitivity between nodes i and j with ultrametric tree (level-based) distance d_{ij} as $\hat{T}_{ij} = \hat{T}^{1/d_{ij}}$. For instance, in **Figures 6**, **7**, two teams sharing the same core cluster on the bottom tree level would have distance $d_{ij} = 1$, two team sin the same cluster on the second-lowest tree level would have distance $d_{ij} = 2$, and so on. We then update α_0 with a recalculated α_{ij} and re-run the Trickling percolation algorithm with a Beta Random field based on Beta random variables from $Be(\alpha_{ij}+1, 1)$. The idea behind this is that Beta random variables from $Be(\alpha_{ij}+1, 1)$ are closer to reality than $Be(\alpha_0+1, 1)$ in the sense that the simulated dominance paths are expected to be expanded and the number of unrealistic simulated "upset" games are significantly reduced. Then a refined version of $\overline{D}(\mathcal{W}_0)$ is computed, and a more efficient estimate $\hat{P}_D(\mathcal{W}_0)$ is constructed. This improvement on estimating $P_D(\mathcal{W}_0)$ turns out to be crucial because of the inherent sensitivity in mimicking directed binary networks as will be discussed in the next subsection.

3.2. Systemic Structures II: Randomness

From the NCAA Football competitive system, the observed directed binary network $\mathcal{G}(\mathcal{W}_t)$ is understood to have two sources

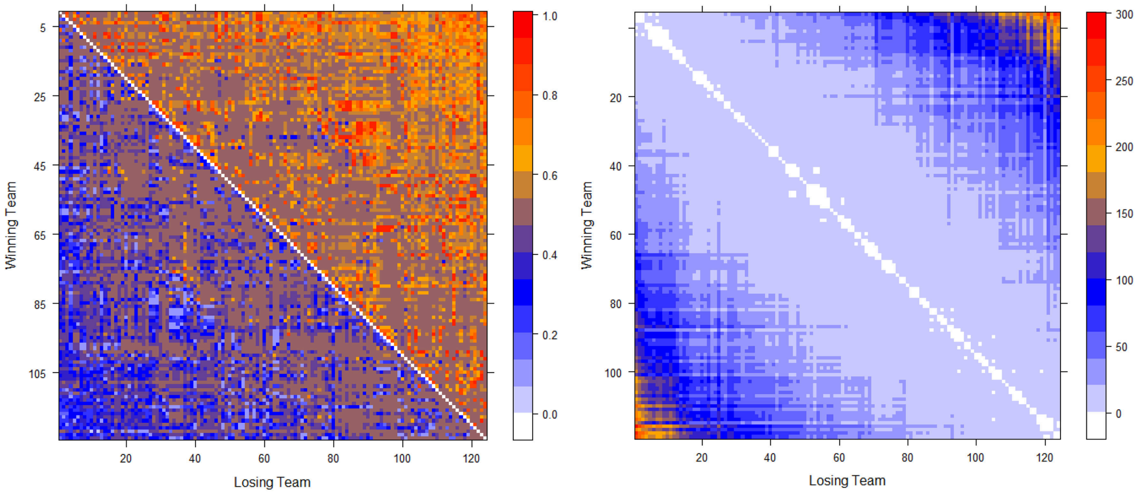

FIGURE 2 | Optimally permuted $\hat{P}_D(\mathcal{W}_0)$ (left panel) and Hamiltonian distance matrix with teams being arranged according to $\hat{\sigma}^*$ for the 2012 NCAA football season (right panel).

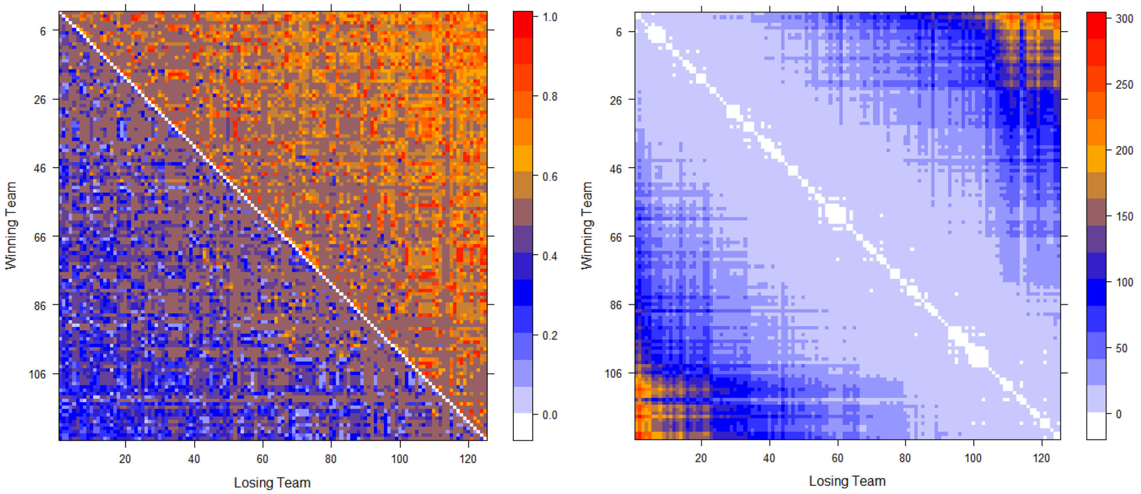

FIGURE 3 | Optimally permuted $\hat{P}_D(\mathcal{W}_0)$ (left panel) and Hamiltonian distance matrix with teams being arranged according to $\hat{\sigma}^*$ for the 2013 NCAA football season (right panel).

of uncertainty: the schedule network $\mathcal{G}(\mathcal{A}_0)$ and the Binomial stochastic mechanism ideally governed by $P_D(\mathcal{W}_0)$, which is responsible for transforming $\mathcal{G}(\mathcal{A}_0)$ into $\mathcal{G}(\mathcal{W}_0)$. We elaborate these sources briefly as follows.

The randomness embedded within a binary undirected network, $\mathcal{A}_0$, has recently been studied in Fushing et al. [16]. It is now understood that such a component of randomness is constrained by its multiscale block patterns organized in the form of a Parisi adjacency matrix.

By taking $\mathcal{G}(\mathcal{A}_0)$ as a thermodynamic system equipped with Ising model potential, this Parisi adjacency matrix is computed as its minimum energy macrostate again by applying the DCG algorithmic computations, see Fushing et al. [16]. Its evident multiscale block patterns are explicitly revealed by superimposing the node organization pertaining to the computed DCG ultrametric tree onto rows and columns of the binary schedule matrix $\mathcal{A}_0$.

The DCG ultrametric trees of the 2012 and 2013 NCAA college football schedules are computed and shown in **Figures 6A**, **7A**, respectively. These Ultrametric trees rediscover the real conference structure of the college football teams. The three singletons are indeed independent teams: Notre Dame, Navy and Air Force. These tree also bring out which conferences are closer, and which are far apart. The resultant Parisi adjacency matrices with their multiscale block patterns are seen in **Figures 2B**, **3B**. Such block information becomes evident when we compare **Figure 1A** with **Figure 6B**, as well as **Figure 1B** with **Figure 7B**. It is noted that the randomness pertaining to any block of $\mathcal{G}(\mathcal{A}_0)$ is the randomness of generating such a sub-matrix by subject to constraints of two

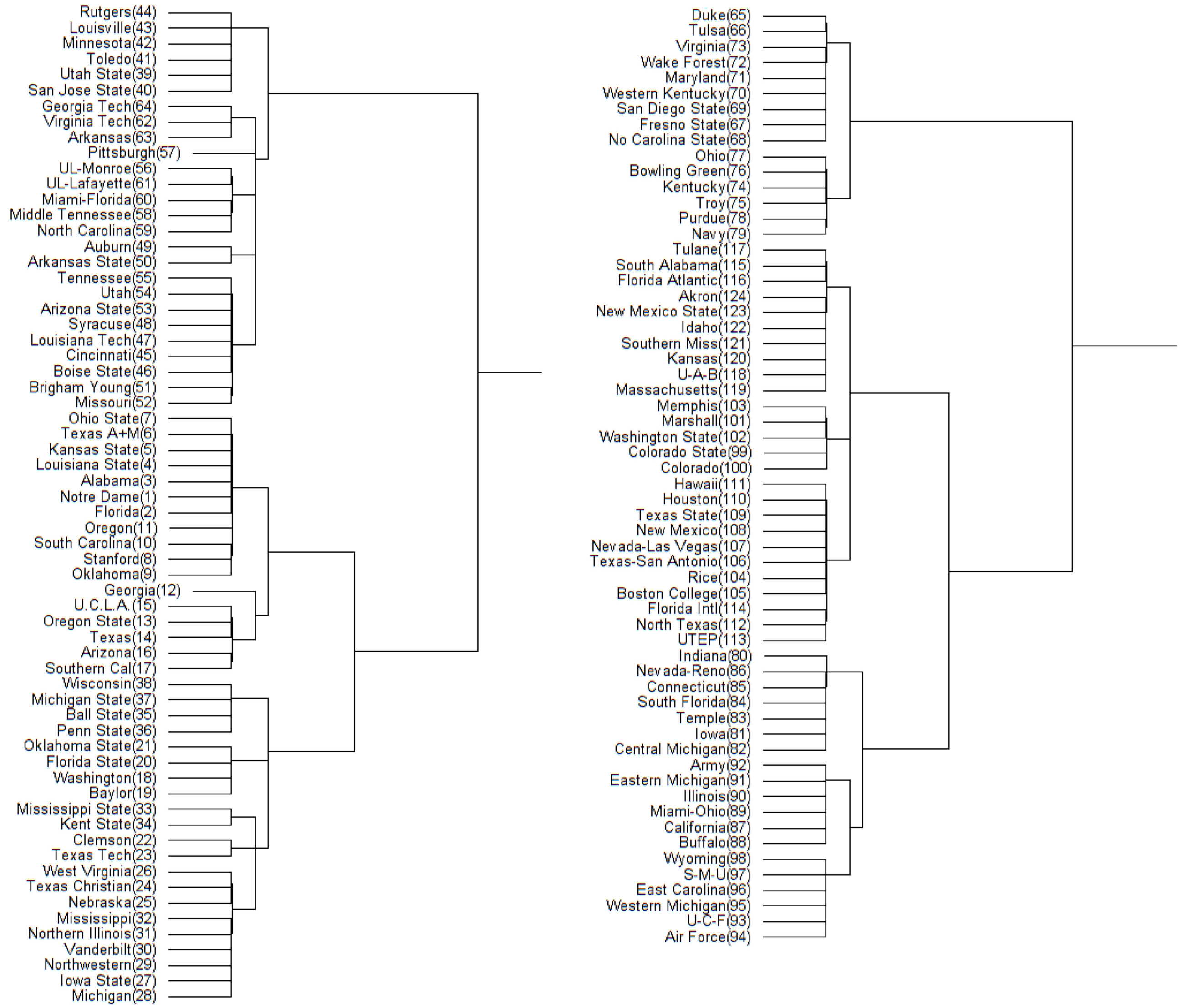

FIGURE 4 | Hamiltonian distance based DCG tree for the 2012 NCAA football season. The tree has been cut at its highest level for ease of presentation. The branch containing the top 64 teams is displayed on left; the remaining teams are displayed on right. The number in the parentheses next to each team name is its position in $\hat{\sigma}^*$.

sequences of row and column sums. Such an explicit use of randomness would be seen in the next section of Network mimicking.

The second source of randomness regarding to transforming $\mathcal{G}(\mathcal{A}_0)$ into $\mathcal{G}(\mathcal{W}_0)$ is pragmatically modeled via a Binomial stochastic mechanism on a game-by-game basis. We assume the global conditional independence given $P_D(\mathcal{W}_0)$ among all football games. This assumption could be realistic under the setting of learning from the network data $\mathcal{G}(\mathcal{W}_0)$. In contrast, such an assumption of independence would be too simplistic when the data set is indeed a sequence of directed binary networks, which is realized along the series of football games sequentially played throughout the entire 13 week season.

3.3. Mimicking Directed Binary Networks and System's Robustness

Mimicking a directed binary network $\mathcal{G}(\mathcal{W}_0)$ is meant to generate a microscopic state, or microstate of the system as one whole. That is to say that any mimicry of directed binary network is required to conform to the computed deterministic structures, while its generating mechanism has to be carried out subject to the constraints pertaining to the identified randomness. An algorithm for mimicking a directed binary network is proposed below. A microstate ensemble can be derived by repeatedly applying this mimicking algorithm. We then make use of such a microstate ensemble to explore the systemic sensitivity, which is typically focused on a noticeable function of this system of interest. Here we designate the power ordering axis on the

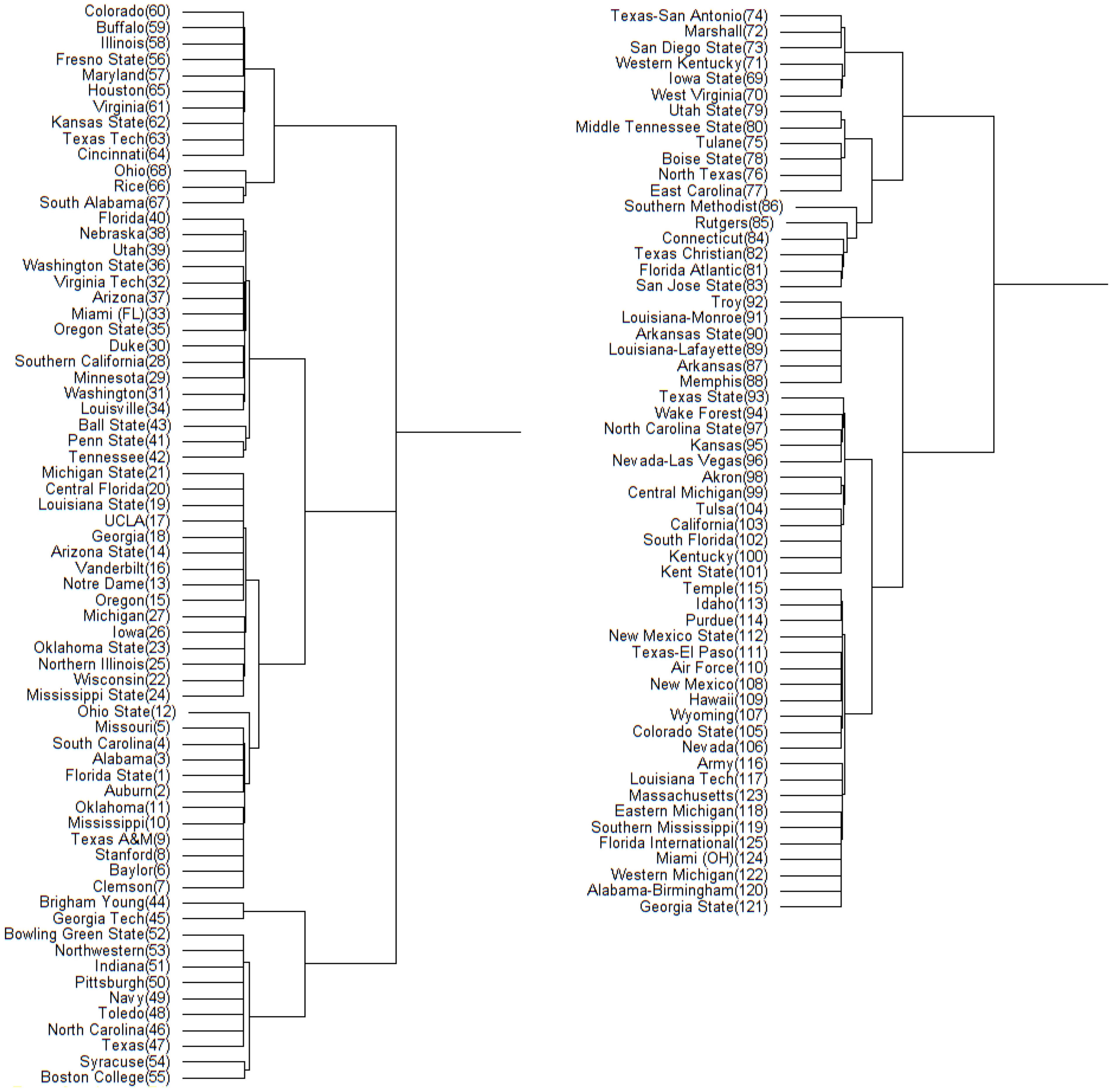

FIGURE 5 | Hamiltonian distance based DCG tree for the 2013 NCAA football season. The tree has been cut at its highest level for ease of presentation. The branch containing the top 68 teams is displayed on left; the remaining teams are displayed on right. The number in the parentheses next to each team name is its position in $\hat{\sigma}^*$.

optimally permuted dominance probability matrix $\hat{P}_D(\mathcal{W}_0$ as the "linear-like ranking function" from top to bottom. Such a linear-like ranking function is then evaluated on each member of the ensemble in the exact same way of computing $\hat{P}_D(\mathcal{W}_0)$. The global variability along the whole spectrum of ranking ordering axis from Number one to Number m is evaluated in this sensitivity investigation.

3.3.1. Directed Binary Network Mimicking Algorithm

NB-1: Simulate a schedule matrix conforming to the finest block-patterns in its Parisi adjacency matrix of $\mathcal{G}(\mathcal{A}_0)$ (see [16] for this computational algorithm);

NB-2: Simulate win-and-loss matrix based upon the simulated schedule matrix and Binomial random mechanism with means being fixed corresponding to $\hat{P}_D(\mathcal{W}_0)$.

By applying the above algorithm, a microstate ensemble of size 100 is generated for each of the two NCAA data on Years 2012 and 2013. Their linear-like ranking functions are evaluated and summarized in **Tables 1**, **2** below. We use the mimicked ranking standard deviation as an index of sensitivity of one team-node. The heterogeneity of sensitivity across the entire power ordering axis is clearly seen. In both tables, there is a noticeable increasing trend in standard deviation as teams become more lowly ranked; standard deviations then decrease again as approaching the most

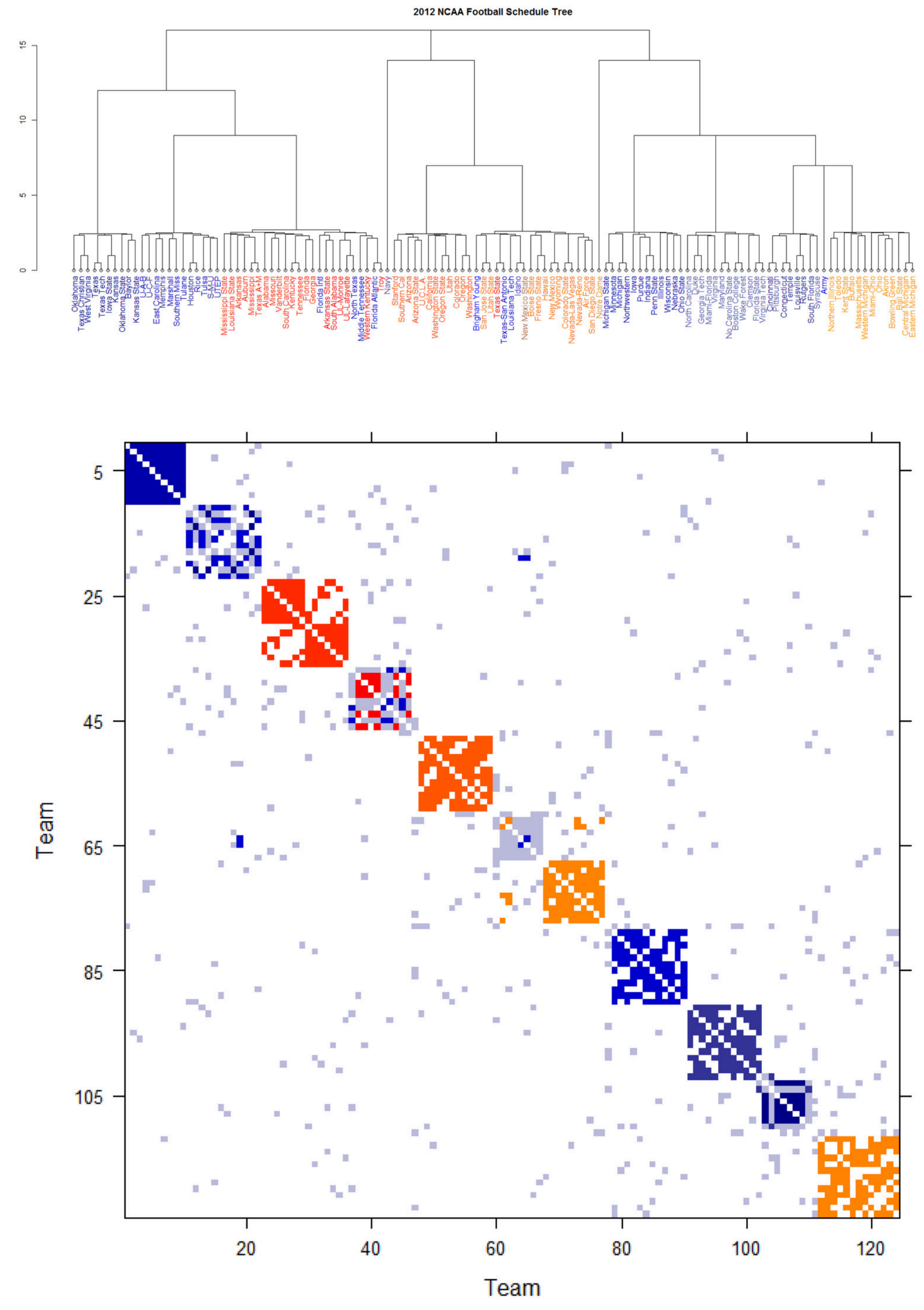

FIGURE 6 | Schedule DCG tree and its induced schedule matrix $\mathcal{S}_0$ of NCAA football data on year 2012: (Upper panel) DCG tree and (Lower panel) Parisi adjacency matrix with colored diagonal blocks marking the intra-conference games for the real conferences, and off-diagonal inter-conference games are marked in gray. The six independent teams are scatted on the tree. Conferences, from blue to red, are ACC, American, Big 12, Big Ten, Conference-USA, Independent schools, Mid-American, Mountain West, Pac 12, SEC, and Sun Belt.

lowly-ranked teams. However, several outliers have significant large standard variations.

In **Table 1**, we notice an outlier, San Jose State, with a significant large standard deviation with respect to its mimicked ranking median. San Jose State had a successful season, only losing two games. However, their conference (Mountain West) is generally considered to be one of the weaker conferences. Since San Jose State plays many of its

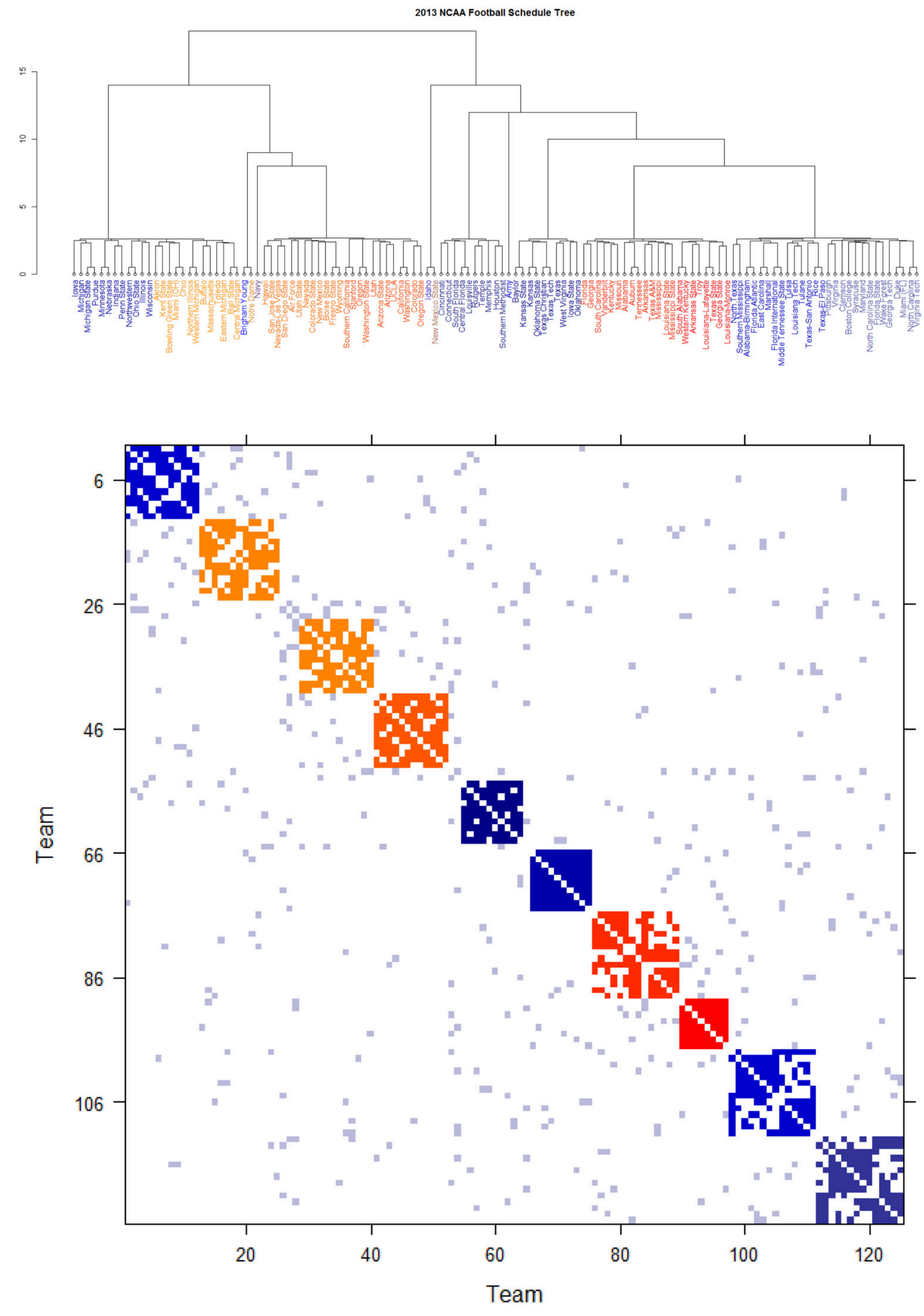

FIGURE 7 | Schedule DCG tree and its induced schedule matrix $\mathcal{S}_0$ of NCAA football data on year 2013: (Upper panel) DCG tree and (Lower panel) Parisi adjacency matrix with colored diagonal blocks marking the intra-conference games for the real conferences, and off-diagonal inter-conference games are marked in gray. The six independent teams are scatted on the tree. Conferences, from blue to red, are ACC, American, Big 12, Big Ten, Conference-USA, Independent schools, Mid-American, Mountain West, Pac 12, SEC, and Sun Belt.

games against teams in the Mountain West, we would expect them to win many of them. While San Jose State's standing within their conference is clear, it is much less apparent where they should be ranked among all teams. Here's a little bit more about San Jose State. It lost to Stanford (ranked 8th) and Utah State (ranked 39th). Our hypothesis is that the simulated annealing wanted to put them as high as possible, without surpassing Utah State. In fact, all 9 teams that San Jose

TABLE 1 | Top 25 NCAA football teams for the 2012 season as ranked with respect to microstate ensemble.

Rank	School	W	L	Conference	Mean	Median	SD
1	Florida	10	1	SEC	4.12	3.0	2.99
2	Alabama	11	1	SEC	5.63	5.0	3.11
3	Notre Dame	12	0	Ind6	5.58	5.0	3.79
4	Stanford	11	2	Pac-12	5.38	5.0	3.12
5	Louisiana State	9	2	SEC	6.03	5.5	3.09
6	Ohio State	12	0	Big Ten	6.82	7.0	3.92
7	Kansas State	10	1	Big 12	7.58	7.5	5.39
8	Georgia	10	2	SEC	7.78	8.0	3.98
9	Texas A&M	8	2	SEC	8.50	8.5	3.97
10	Oregon	10	1	Pac-12	8.61	9.0	3.31
11	South Carolina	9	2	SEC	9.71	10.0	3.51
12	San Jose State	9	2	Mountain West	18.13	13.5	14.40
13	Oregon State	8	3	Pac-12	13.67	14.0	2.53
14	Oklahoma	9	2	Big 12	14.45	15.0	5.94
15	UCLA	9	4	Pac-12	16.33	16.0	3.26
16	Arizona	6	5	Pac-12	19.18	18.0	4.10
17	Southern Cal	7	5	Pac-12	19.66	18.0	4.73
18	Nebraska	9	3	Big Ten	21.96	20.0	6.10
19	Arizona State	6	5	Pac-12	21.55	20.5	5.63
20	Michigan	8	4	Big Ten	23.89	22.0	8.07
21	Penn State	8	4	Big Ten	27.02	23.5	13.82
22	Texas	8	4	Big 12	23.54	23.5	4.24
23	Vanderbilt	7	4	SEC	26.41	24.0	7.76
24	Missouri	4	7	SEC	27.48	25.0	10.71
25	Northern Illinois	11	1	MAC	25.11	25.0	16.29

Mean, median, and standard deviation of ranks from 100 mimicking iterations are reported. The first column "Rank" is primarily based on the ranking on the 7th column "Median." When there is a tie, the ranking is resolved by comparing the "winning" proportion among the 100 mimicking iterations.

TABLE 2 | Top 25 NCAA football teams for the 2013 season as ranked with respect to microstate ensemble.

Rank	School	W	L	Conference	Mean	Median	SD
1	Auburn	11	1	SEC	3.81	3.0	2.57
2	Florida State	12	0	ACC	4.09	3.0	4.84
3	Alabama	10	1	SEC	4.35	4.0	2.56
4	Michigan State	11	1	Big Ten	9.41	6.0	8.64
5	Missouri	10	2	SEC	8.59	8.0	4.16
6	Oklahoma State	9	2	Big 12	10.94	8.0	9.63
7	Ohio State	11	1	Big Ten	12.09	9.0	8.66
8	Stanford	11	2	Pac-12	9.19	9.0	5.23
9	South Carolina	9	2	SEC	10.43	10.0	4.62
10	Arizona State	9	3	Pac-12	12.84	12.0	6.03
11	Baylor	10	1	Big 12	15.45	12.0	9.44
12	Clemson	8	2	ACC	14.26	12.0	9.06
13	Georgia	7	4	SEC	14.39	13.0	6.66
14	Central Florida	11	1	American	15.10	15.0	10.78
15	Louisiana State	8	3	SEC	15.35	15.0	6.25
16	Notre Dame	8	4	Ind6	15.69	16.0	8.52
17	Mississippi	6	5	SEC	18.02	16.5	8.75
18	Texas A&M	7	4	SEC	17.16	17.0	5.65
19	Oregon	9	2	Pac-12	18.42	18.0	7.09
20	UCLA	9	3	Pac-12	19.66	19.5	5.86
21	Vanderbilt	7	4	SEC	23.62	23.0	7.33
22	Southern Cal	9	4	Pac-12	24.21	24.0	6.48
23	Oklahoma	10	2	Big 12	27.37	25.0	11.03
24	Texas	8	4	Big 12	27.95	25.5	11.41
25	Washington	7	4	Pac-12	26.35	26.0	6.33

Mean, median, and standard deviation of ranks from 100 mimicking iterations are reported. The first column "Rank" is primarily based on the ranking on the 7th column "Median." When there is a tie, the ranking is resolved by comparing the "winning" proportion among the 100 mimicking iterations.

State defeated are ranked below them, so with respect to only their own schedule, they are ranked appropriately at 40 in the computed $\hat{\sigma}^*$. Why is it ranked 12th based on the microstate ensemble?

We also notice from their schedule that it played 4 independent teams. These games are necessarily played again in every simulated schedule, since each independent team is treated as its own conference. Since San Jose State won all 4 of these games, they will have a very good chance of winning those 4 games again in the mimicked seasons, since those games were actually played, the dominance probability is particularly high. However, it might not play with the Utah State in mimicked seasons. If they meet, then San Jose State could even win over Utah State with a positive probability. Due to such uncertainty in the mimicked schedule and in Beta random field adapted to $\hat{P}_D(\mathcal{W}_0)$, San Jose State has a significant large variation in linear ranking. We see this inflation of standard deviation for top teams in weaker conferences (Northern Illinois in 2012 and Central Florida in 2013 have the same characteristic), and also for bottom-dwelling teams in strong conferences. This phenomenon to some extent nicely explains the sensitivity of this network mimicking.

We also see that the microstate ensemble does not always place undefeated teams at the top of the rankings. In **Table 1**, Notre Dame and Ohio State were both undefeated, but were ranked third and sixth, respectively. While we do not directly observe evidence for any other teams to be more highly ranked, we must accept that neither team is from the most dominant conference in 2012, the SEC. Neither team had many games played against other highly-ranked teams, while SEC teams such as Florida and Alabama had a higher level of competition. In 2013, the same phenomenon occurs with Florida State. Florida State's conference, the ACC, is broadly considered a weaker conference than the SEC. As a result, we are more hesitant to give Florida State the top ranking because their level of competition was lower than Auburn's, even though Auburn lost one game.

In summary, aforementioned episodes linking to the heterogeneity in ranking sensitivity clearly indicates that NCAA college football is a rather complex system of competition.

As a final note on our sensitivity exploration, beyond our mimicking algorithm, other simulated ensembles of $\mathcal{G}(\mathcal{W}_0)$ could be derived, for instance, by enlarging focal scales on the Parisi adjacency matrix on $\mathcal{G}(\mathcal{A}_0)$. With the NCAA data, this type of simulated ensemble corresponds to schedules swapping between different conferences. Our experimental results drastically vary from what are reported here in the **Tables 1**, **2** even when schedule swapping is only limited between nearest conferences. This sensitivity is likely due to the fact that the dyadic data in NCAA data is sparse. Consider a simulated game involving two teams in two different conferences. The two teams likely have few common opponents. Any upset loss resulted from such a simulated game would create a large amount of uncertainty when it comes to predicting the more dominant team. Even such a seemingly unremarkable simulated event may obscure the whole power structure.

4. DISCUSSION

Our computational and algorithmic developments on directed binary networks show the capability of extracting deterministic structures and randomness of the underlying complex system of interest. The compositional constructs advance our systemic understanding in a critical way. The platform of network mimicking is proven to have impacts on exploring a system's functional sensitivity. By putting together all computed results from all considered perspectives, we should be able to advance further our understanding of a complex system.

Through the illustrating NCAA college football example, we clearly see complexity embedded within this competitive system, while at the same time we are able to recognize realistically the complicated information contents contained in a seemingly simplistic directed binary network. This fact reminds us that understanding a system of scientific interest does require a heavy investment on computing.

At this stage there are still many key technical difficulties remaining unsolved. They are tentatively resolved with heuristic approaches or proposals in this paper. Some questions that particularly need rigorous endeavors are listed as follows: How can we effectively measure and estimate pairwise transitivity T_{ij} and the related parameter α_{ij}? How much confidence can be placed on the blank entries of $\hat{P}_D(\mathcal{W}_0)$?

Here it is worth emphasizing again the network mimicking principle: any mimicked network must be taken as a microscopic state that is generated coherently with respect to the system's randomness, and conforms to the system's deterministic structures.

The proposed directed binary network mimicking algorithm would allow us to evaluate the network entropy. This entropy should be the sum of the entropy pertaining to the finest scale block patterns of the Parisi adjacency matrix (which is computed by measuring the sizes of corresponding microstate ensembles as discussed in [16]), plus the entropy derived from the Binomial random mechanism with means being fixed corresponding to $\hat{P}_D(\mathcal{W}_0)$.

On the other hand. this NCAA football analysis comes at a particularly opportune time, as the NCAA recently changed the postseason format from a single championship game to a four-team playoff. The teams participating in the playoff will be decided by an appointed committee, rather than by a computer and/or voting system like the Bowl Championship Series (BCS). A linear-like ranking hierarchy such as the one developed in this paper could be applied to provide the committee with some objectivity in choosing the participating schools, or even the number of schools that are to participate. Trees such as those in **Figures 4**, **5** can be used to determine the contents of the highest tier of football teams, while rankings derived based on the microstate ensemble, holding conferences constant, provide insight into the level of uncertainty we might have about teams' relative standing.

As a final remark, we mention one implication of our sensitivity explorations on the power hierarchy of NCAA Football League. These investigations clearly show the difficulties in pertinent and realistic network modeling and its statistical inferences. At this stage there is no available methodology or even knowledge for how to physically generate multiscale block-patterns without knowing the tree structure priori, neither for the power structural hierarchy. The unavailability is not surprising because that a network is better perceived as a complex system itself. The network mimicking principle proposed here might serve as a guide on network modeling, goodness-of-fit testing and making inferences.

AUTHOR CONTRIBUTIONS

FH and KF designed the experiments and analyzed data . FH did most writing and KF performed all computing.

ACKNOWLEDGMENTS

This research is supported in part by the NSF under Grants No: DMS 1007219 [co-funded by Cyber-enabled Discovery and Innovation (CDI) program].

REFERENCES

1. Crutchfield JP. Between order and chaos. *Nat Phys.* (2012) **8**:17–24. doi: 10.1038/nphys2190
2. Simon HA. The architecture of complexity. *Proc Am Philos Soc.* (1962) **106**:467–82.
3. Anderson PW. More is different. *Science* (1972) **177**:393–6. doi: 10.1126/science.177.4047.393
4. Crutchfield JP, Machta J. Introduction to focus issue on "randomness, structure, and causality: measure of complexity from theory to applications. *Chaos* (2011) **21**:037101.
5. Albert R, Barabási AL. Statistical mechanics of complex network. *Rev Mod Phys.* (2002) **74:**47–97.
6. Barzel B, Barabási AL. Universality in network dynamics. *Nat Phys.* (2013) **9:**673–81.
7. Gao J, Barzel B, Barabási AL. Universal resilience patterns in complex networks. *Nature* (2016) **530:**307–12. doi: 10.1038/nature18019
8. Doyle JC, Csete M. Architecture, constraints and behavior. *Proc Natl Acad Sci USA* (2011) **108**:15624–30. doi: 10.1073/pnas.1103557108
9. Fujii K, Fushing H, Beisner B, McCowan B. *Computing Power Structures in Directed Biosocial Networks: Flow Percolation and Imputed Conductance*. Technical Report. Department of Statistics, UC Davis, Davis (2016).

10. Landau HG. On dominance relations and the structure of animal societies. I. effect of inherent characters. *Bull Math Biophys.* (1951) **13**:1–19. doi: 10.1007/BF02478336
11. Kendall MG. *Rank Correlation Methods.* London: Charles Griffin (1962).
12. Appleby MC. The probability of linearity in hierarchies. *Anim Behav.* (1983) **31**:600–8. doi: 10.1016/S0003-3472(83)80084-0
13. Kasuya E. A randomization test for linearity of dominance hierarchies. J Ethol. (1995) **13**:137–40. doi: 10.1007/BF02352574
14. de Vries H. Finding a dominance order most consistent with a linear hierarchy: a new procedure and review. Anim Behav. (1998) **55**:827–43. doi: 10.1006/anbe.1997.0708
15. Leicht EA, Newman MEJ. Community structure in directed networks. *Phys Rev Lett.* (2008) **100**:118703. doi: 10.1103/PhysRevLett.100.118703
16. Fushing H, Chen C, Liu S, Koehl P. Bootstrapping on undirected binary networks via statistical mechanics. *J Stat Phys.* (2014) **156**:823–42. doi: 10.1007/s10955-014-1043-6
17. Fushing H, McAssey MP, McCowan B. Computing a ranking network with confidence bounds from a graph-based Beta random field. *Proc R Soc A* (2011) **467**:3590–612. doi: 10.1098/rspa.2011.0268
18. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science* (1983) **220**:671–80. doi: 10.1126/science.220.4598.671
19. Fushing H, McAssey MP. Time, temperature and data cloud geometry. *Phys Rev E* (2010) **82**:061110. doi: 10.1103/PhysRevE.82.061110
20. Fushing H, Wang H, Van der Waal K, McCowan B, Koehl P. Multi-scale clustering by buildi ng a robust and self-correcting ultrametric topology on data points. *PLOS ONE* (2013) **8**:e56259. doi: 10.1371/journal.pone.0056259

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.