Power in Normative Systems

Thomas Ågotnes[†] Wiebe van der Hoek[‡] Moshe Tennenholtz[*] Michael Wooldridge[‡]

[†]Computer Engineering
Bergen University College
Bergen, Norway
tag@hib.no

[‡]Computer Science
University of Liverpool
Liverpool, UK
{wiebe,mjw}@csc.liv.ac.uk

[*]Microsoft Israel R&D Center &
Technion–Israel Institute of Technology
Israel
moshet@ie.technion.ac.il

ABSTRACT

Power indices such as the Banzhaf index were originally developed within voting theory in an attempt to rigorously characterise the influence that a voter is able to wield in a particular voting game. In this paper, we show how such power indices can be applied to understanding the relative importance of agents when we attempt to devise a coordination mechanism using the paradigm of social laws, or normative systems. Understanding how pivotal an agent is with respect to the success of a particular social law is of benefit when designing such social laws: we might typically aim to ensure that power is distributed evenly amongst the agents in a system, to avoid bottlenecks or single points of failure. After formally defining the framework and illustrating the role of power indices in it, we investigate the complexity of computing these indices, showing that the characteristic complexity result is #P-completeness. We then investigate cases where computing indices is computationally easy.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems; I.2.4 [**Knowledge representation formalisms and methods**]

General Terms

Theory

Keywords

normative systems, logic, coalitional games, complexity

1. INTRODUCTION

Normative systems (a.k.a. social laws) have been widely promoted as an approach to coordinating multi-agent systems [12, 13, 14, 1]. The idea is that a normative system is a set of prohibitions on the behaviour of agents in a system; after imposing these prohibitions, the designer of the normative system intends that some desirable overall objective will hold.

One of the most important issues associated with normative systems is that of *compliance*: what happens if some agents do not comply with the prohibitions of the normative system? It seems inevitable that non-compliance will occur in real systems, either deliberately (for example, if some participant believes non-compliance is in its interests), or accidentally (for example, as the result of a system crash). It makes sense, therefore, for the designer of a normative system to take into account the possibility of non-compliance at design time. It might be possible to design a normative system so that compliance is the rational choice for participants [1]. However, this approach does not help with the issue of accidental non-compliance (or deliberately irrational behaviour) and it is therefore to the issue of non-compliance, irrespective of its causes, that we address ourselves in the present paper.

The key idea of the paper is to develop principled techniques for measuring the *influence* or *power* that a participant agent has with respect to the success or otherwise of a particular normative system. The approach we adopt makes use of *voting power indices* [8]. Power indices, such as the Banzhaf score, Banzhaf measure, Banzhaf index, and Shapley-Shubik index were originally developed within voting theory in an attempt to rigorously characterise the influence that a voter is able to wield in a particular voting game. In our setting, power is interpreted as the ability of an agent to affect whether or not a normative system has the desired effect. An agent wields such power by choosing to comply or not comply with the prohibitions of the normative system. We would typically aim to ensure that power is distributed *evenly* amongst the agents in a system, so as to avoid bottlenecks or single points of failure. However, we believe the approach also has wider value as an analytical tool, enabling a designer to understand where the key risks or vulnerabilities in a normative system lie. For example, we might use the power distribution to guide the allocation of a maintenance budget, focusing the budget on those participants with a high power index, and hence whose failure to comply would likely be particularly damaging (cf. [3]).

After formally defining the framework of normative systems, we show how power indices can be interpreted within it, and give a detailed example to illustrate their use. We then investigate the computational complexity of computing power indices, showing that the characteristic complexity result is #P-completeness. More precisely, we show that the problem of computing the Banzhaf score is #P-complete, while computing the Banzhaf measure, Banzhaf index, and Shapley-Shubik index are #P-equivalent. We investigate a number of related computational problems, and then investigate cases where computing indices is computationally easy.

Cite as: Power in Normative Systems, Thomas Ågotnes, Wiebe van der Hoek, Moshe Tennenholtz, Michael Wooldridge, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009),* Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. 145–152

2. NORMATIVE SYSTEMS

We use the framework of [14, 1], which uses Kripke structures to model systems, and the logic CTL to characterise the desirable properties of normative systems.

Kripke Structures: We use *Kripke structures* as our basic semantic model for multi-agent systems [7]. A Kripke structure is essentially a directed graph, with a vertex set S corresponding to possible *states* of the system being modelled, and a relation $R \subseteq S \times S$ capturing the possible *transitions* of the system; $S^0 \subseteq S$ denotes the *initial states* of the system. Intuitively, transitions are caused by *agents* in the system performing *actions*, although we do not include such actions in our semantic model (see, e.g., [12, 14] for models which include actions as first class citizens). An arc $(s, s') \in R$ corresponds to the execution of an atomic action by one of the agents in the system. Note that we are therefore here *not* modelling *synchronous* action. This assumption is not essential, but it simplifies the presentation. However, we find it convenient to include within our model the agents that cause transitions. We therefore assume a set A of agents, and we label each transition in R with the agent that causes the transition via a function $\alpha : R \to A$. Finally, we use a vocabulary $\Phi = \{p, q, \ldots\}$ of Boolean variables to express the properties of individual states S: we use a function $V : S \to 2^\Phi$ to label each state with the Boolean variables true (or satisfied) in that state.

Formally, an *agent-labelled Kripke structure* (over Φ) is a 6-tuple $K = \langle S, S^0, R, A, \alpha, V\rangle$, where: S is a finite, non-empty set of states; $S^0 \subseteq S$ ($S^0 \neq \emptyset$) is the set of initial states; $R \subseteq S \times S$ is a total binary transition relation on S; $A = \{1, \ldots, n\}$ is a set of agents; $\alpha : R \to A$ labels each transition in R with an agent; and $V : S \to 2^\Phi$ labels each state with the set of propositional variables true in that state. We hereafter refer to an agent-labelled Kripke structure simply as a *Kripke structure.*

A *path* over a transition relation R is an infinite sequence of states $\pi = s_0, s_1, \ldots$ such that $\forall u \in \mathbb{N}$: $(s_u, s_{u+1}) \in R$. If $u \in \mathbb{N}$, then we denote by $\pi[u]$ the component indexed by u in π (thus $\pi[0]$ denotes the first element, $\pi[1]$ the second, and so on). A path π such that $\pi[0] = s$ is an *s-path.* Let $\Pi_R(s)$ denote the set of s-paths over R; since it will usually be clear from context, we often omit reference to R, and simply write $\Pi(s)$.

CTL: To express the *objectives* of normative systems, we use Computation Tree Logic (CTL), a well-known and widely used branching time temporal logic [7]. Given a set $\Phi = \{p, q, \ldots\}$ of atomic propositions, the syntax of CTL is defined by the following grammar, where $p \in \Phi$:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathsf{E}\bigcirc\varphi \mid \mathsf{E}(\varphi\,\mathcal{U}\,\varphi) \mid \mathsf{A}\bigcirc\varphi \mid \mathsf{A}(\varphi\,\mathcal{U}\,\varphi)$$

The semantics of CTL are given with respect to the satisfaction relation "$\models$", which holds between *pointed structures* of the form K, s (where K is a Kripke structure and s is a state in K), and formulae of the language. The satisfaction relation is defined as follows:

$K, s \models \top$;

$K, s \models p$ iff $p \in V(s)$ (where $p \in \Phi$);

$K, s \models \neg\varphi$ iff not $K, s \models \varphi$;

$K, s \models \varphi \vee \psi$ iff $K, s \models \varphi$ or $K, s \models \psi$;

$K, s \models \mathsf{A}\bigcirc\varphi$ iff $\forall\pi \in \Pi(s) : K, \pi[1] \models \varphi$;

$K, s \models \mathsf{E}\bigcirc\varphi$ iff $\exists\pi \in \Pi(s) : K, \pi[1] \models \varphi$;

$K, s \models \mathsf{A}(\varphi\,\mathcal{U}\,\psi)$ iff $\forall\pi \in \Pi(s), \exists u \in \mathbb{N}$, s.t. $K, \pi[u] \models \psi$ and $\forall v, (0 \leq v < u) : K, \pi[v] \models \varphi$

$K, s \models \mathsf{E}(\varphi\,\mathcal{U}\,\psi)$ iff $\exists\pi \in \Pi(s), \exists u \in \mathbb{N}$, s.t. $K, \pi[u] \models \psi$ and $\forall v, (0 \leq v < u) : K, \pi[v] \models \varphi$

The remaining classical logic connectives ("$\wedge$", "$\rightarrow$", "$\leftrightarrow$") are defined as abbreviations in terms of $\neg, \vee$ in the conventional way. The remaining CTL temporal operators are defined as follows:

$$\begin{array}{lcl@{\qquad}lcl} \mathsf{A}\Diamond\varphi & \equiv & \mathsf{A}(\top\,\mathcal{U}\,\varphi) & \mathsf{E}\Diamond\varphi & \equiv & \mathsf{E}(\top\,\mathcal{U}\,\varphi) \\ \mathsf{A}\,\Box\varphi & \equiv & \neg\mathsf{E}\Diamond\neg\varphi & \mathsf{E}\,\Box\varphi & \equiv & \neg\mathsf{A}\Diamond\neg\varphi \end{array}$$

The problem of checking whether $K, s \models \varphi$ for given K, s, φ (*model checking*) can be done in deterministic polynomial time [7]. We write $K \models \varphi$ if $K, s_0 \models \varphi$ for all $s_0 \in S^0$.

Later, we use two fragments of CTL: the universal language L^u (with typical element u), and the existential fragment L^e (typical element ε):

$$\begin{array}{l} u ::= \top \mid \bot \mid p \mid \neg p \mid u \vee u \mid u \wedge u \mid \mathsf{A}\bigcirc u \mid \mathsf{A}\,\Box u \mid \mathsf{A}(u\,\mathcal{U}\,u) \\ \varepsilon ::= \top \mid \bot \mid p \mid \neg p \mid \varepsilon \vee \varepsilon \mid \varepsilon \wedge \varepsilon \mid \mathsf{E}\bigcirc\varepsilon \mid \mathsf{E}\,\Box\varepsilon \mid \mathsf{E}(\varepsilon\,\mathcal{U}\,\varepsilon) \end{array}$$

The key point about these fragments is as follows. Let us say, for two Kripke structures $K_1 = \langle S, S^0, R_1, A, \alpha, V\rangle$ and $K_2 = \langle S, S^0, R_2, A, \alpha, V\rangle$ that K_1 is a subsystem of K_2 and K_2 is a supersystem of K_1, (denoted $K_1 \sqsubseteq K_2$), iff $R_1 \subseteq R_2$. Then we have:

THEOREM 1 ([14]). *Suppose $K_1 \sqsubseteq K_2$, and $s \in S$. Then:*

$$\begin{array}{lcl} \forall\varepsilon \in L^e : K_1, s \models \varepsilon & \Rightarrow & K_2, s \models \varepsilon; \quad \textit{and} \\ \forall u \in L^u : K_2, s \models u & \Rightarrow & K_1, s \models u. \end{array}$$

Normative Systems: For our purposes, a *normative system* (or "norm") is simply *a set of constraints on the behaviour of agents in a system.* Formally, a normative system η (w.r.t. a Kripke structure $K = \langle S, S^0, R, A, \alpha, V\rangle$) is simply a subset of R, such that $R \setminus \eta$ is a total relation (i.e., every state has a successor: for every $s \in S$ there is a $t \in S$ such that $(s, t) \in R$), with the intended interpretation that *the transitions in η are forbidden.* The requirement that $R \setminus \eta$ is total is a *reasonableness* constraint: it prevents normative systems which lead to states with no successor. (This assumption allows us to use CTL as the object language. It is no limitation, in the sense that a system being 'stuck' can be modelled as 'looping in the same state forever'). Let $N(R) = \{\eta \subseteq R : R \setminus \eta \text{ is total}\}$ be the set of normative systems over R. Let $A(\eta) = \{\alpha(s, s') : (s, s') \in \eta\}$ denote the set of agents involved in η.

The effect of *implementing* a normative system on a Kripke structure is to eliminate from it all transitions that are forbidden according to this normative system (see [14] and, for an approach that incentivises agents to keep the norm, [1]). If K is a Kripke structure, and η is a normative system over K, then $K \dagger \eta$ denotes the Kripke structure obtained from K by deleting transitions forbidden in η. Formally, if $K = \langle S, S^0, R, A, \alpha, V\rangle$, and $\eta \in N(R)$, then let $K \dagger \eta = K'$ be the Kripke structure $K' = \langle S', S^{0\prime}, R', A', \alpha', V'\rangle$ where:

- $S = S'$, $S^0 = S^{0\prime}$, $A = A'$, and $V = V'$;
- $R' = R \setminus \eta$; and
- α' is the restriction of α to R'.

The most basic question we can ask in the context of normative systems is as follows. We are given a Kripke structure K, representing the state transition graph of our system, and we are given a CTL formula φ, representing the *objective* of a normative system designer (that is, the objective characterises what a designer wishes to accomplish with a normative system). The *feasibility* problem is then whether or not there exists a normative system η such that implementing η in K will achieve φ, i.e., whether $K \dagger \eta \models \varphi$. In general, given a Kripke structure K and CTL objective φ, checking feasibility is NP-complete [12, 14]. We say that η is *effective* for φ in K if $K \dagger \eta \models \varphi$. Let $\mathit{eff}(K, \varphi)$ denote the set of normative systems that are effective for φ in K:

$$\mathit{eff}(K, \varphi) = \{\eta \in N(R) : K \dagger \eta \models \varphi\}.$$

A *social system* $S = \langle K, \varphi, \eta \rangle$ consists of a Kripke structure K, representing the dynamics of the system, a CTL formula φ, representing the objective that a designer has, and a normative system η, by which means the designer intends to achieve the objective.

We make use of operators on normative systems which correspond to groups of agents "defecting" from the normative system. Formally, let $K = \langle S, S^0, R, A, \alpha, V \rangle$ be a Kripke structure, let $C \subseteq A$ be a set of agents over K, and let η be a normative system over K. Then $\eta \upharpoonright C$ denotes the normative system that is the same as η except that it only contains the arcs of η that correspond to the actions of agents in C, i.e., $\eta \upharpoonright C = \{(s, s') \in \eta : \alpha(s, s') \in C\}$.

3. COALITIONAL GAMES AND POWER

We need some definitions from the area of cooperative game theory [11] and the theory of voting power [8]. A *cooperative* (or *coalitional*) *game* is a pair $G = \langle A, \nu \rangle$, where $A = \{1, \ldots, n\}$ is a set of *players*, or *agents*, and $\nu : 2^A \to \mathbb{R}$ is the *characteristic function* of the game, which assigns to every set of agents a numeric value, intuitively corresponding to the utility that this group of agents could obtain if they chose to cooperate. Notice that this model does not attempt to model *how* groups of agents might cooperate, or *where* utility comes from; nor does it dictate how the utility obtained by a group of agents should be distributed among coalition members. A cooperative game is said to be *simple* if the range of ν is $\{0, 1\}$; in simple games we say that C are *winning* if $\nu(C) = 1$, while if $\nu(C) = 0$, we say C are *losing*. For simple games, a number of *power indices* attempt to characterise in a systematic way the *influence* that a given agent has, by measuring how effective this agent is at turning a losing coalition into a winning coalition [8]. The best-known of these is perhaps the *Banhzaf index* and its relatives, the Banzhaf score and Banzhaf measure [4].

Agent i is said to be a *swing player* for $C \subseteq A \setminus \{i\}$ if C is not winning but $C \cup \{i\}$ is. We define a function $\mathit{swing}(C, i)$ (where $i \notin C$) so that this function returns 1 if i is a swing player for C, and 0 otherwise, i.e.,

$$\mathit{swing}(C, i) = \begin{cases} 1 & \text{if } \nu(C) = 0 \text{ and } \nu(C \cup \{i\}) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Now, we define the *Banzhaf score* for agent i, denoted σ_i, to be the number of coalitions for which i is a swing player [8, p.39]:

$$\sigma_i = \sum_{C \subseteq A \setminus \{i\}} \mathit{swing}(C, i). \tag{1}$$

The *Banzhaf measure*, denoted μ_i, is the probability that i would be a swing player for a coalition chosen at random from $2^{A \setminus \{i\}}$ [8, p.39]:

$$\mu_i = \frac{\sigma_i}{2^{n-1}} \tag{2}$$

The *Banzhaf index* for a player $i \in A$, denoted by β_i, is the proportion of coalitions for which i is a swing to the total number of swings in the game – thus the Banzhaf index is a measure of relative power, since it takes into account the Banzhaf score of other agents [8, p.39]:

$$\beta_i = \frac{\sigma_i}{\sum_{j \in A} \sigma_j} \tag{3}$$

Finally, we define the *Shapley-Shubik index* [8, p.39]; here the *order* in which agents join a coalition plays a role. Let $P(A)$ denote the set of all permutations of A, with typical members ϖ, ϖ', etc. If $\varpi \in P(A)$ and $i \in A$, then let $\mathit{prec}(i, \varpi)$ denote the members of A that precede i in the ordering ϖ. (For example, if $\varpi = (a_3, a_1, a_2)$, then $\mathit{prec}(a_2, \varpi) = \{a_1, a_3\}$.) Given this, let ς_i denote the Shapley-Shubik index of i, defined as follows [8, p.196]:

$$\varsigma_i = \frac{1}{|A|!} \sum_{\varpi \in P(A)} \mathit{swing}(\mathit{prec}(i, \varpi), i) \tag{4}$$

Thus the Shapley-Shubik index is essentially the Shapley value [11, p.291] applied to simple ($\{0, 1\}$-valued) cooperative games. See also Example 1.

4. POWER IN SOCIAL SYSTEMS

We now make our link between, on the one hand, normative systems and the issue of compliance and, on the other hand, cooperative games and power indices. The idea is as follows. Suppose we are given a social system $S = \langle K, \varphi, \eta \rangle$, i.e., a Kripke structure representing a system, a CTL formula representing the objective that the designer wishes to accomplish, and a normative system η, by which means the designer wishes to accomplish φ. Now, it seems very natural that the designer of the normative system would want to consider *how important* the various agents within the system are with respect to the correct operation of the normative system. Our aim is to use the power metrics discussed above for this purpose.

To use power indices in normative systems, we must first show how to associate a coalitional game with a social system. The intuition is that a value of 1 is assigned to a coalition C if C complying with the normative system will achieve the objective, and 0 otherwise. Formally, a social system $S = \langle K, \varphi, \eta \rangle$ (where $K = \langle S, S^0, R, A, \alpha, V \rangle$), induces a simple cooperative game $G(S) = \langle A, \nu_S \rangle$, where the set of players A is as in K, and ν_S is defined as follows:

$$\nu_S(C) = \begin{cases} 1 & \text{if } K \dagger (\eta \upharpoonright C) \models \varphi \\ 0 & \text{otherwise.} \end{cases}$$

We can then directly apply the indices discussed above to understand the relative power that agents have in a social system. Power, in this sense, will mean the relative ability of an agent to cause a normative system to succeed or fail with respect to the objective. The reason for wanting to measure power in this way is not machiavellian: it at least *identifies* agents that are crucial in achieving the objective, and one might desire to ensure that power is distributed *as*

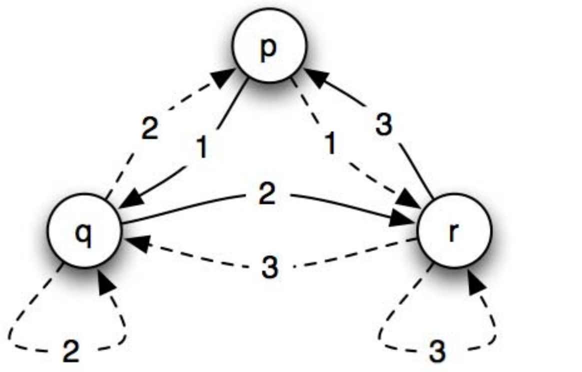

Figure 1: Kripke model M, representing a resource-passing scenario.

evenly as possible, in order to ensure that there are are no bottlenecks, or single-points-of-failure.

Note that one could, of course, alternatively measure the power an agent has to ensure the achievement of an objective by *defecting* from a normative system, i.e., by *not* obeying the norm. However, we will restrict our attention to power exerted by compliance. This matches our intuitions about compliance, where we think of the more agents complying, the better. This property is captured in the idea of *coalition monotonicity*, ensuring that a compliant coalition never have to fear new members joining them:

$$\forall C : K \dagger (\eta \upharpoonright C) \models \varphi \text{ implies } \forall C' \supseteq C : K \dagger (\eta \upharpoonright C') \models \varphi.$$

Some types of social system are inherently monotone in this sense.

Proposition 1. *If $\varphi \in L^u$ then the social system $\langle K, \varphi, \eta \rangle$ is coalition monotone.*

Note that we do *not* in general assume that social systems are coalition monotone in this paper.

Let us consider an example of our power measures.

Example 1 (Passing on a resource). *Figure 1 shows a simple example of a Kripke structure. Here, we have $A = \{1, 2, 3\}$, and the idea is that the agents can pass a resource to each other. Each state is labelled with an atom, indicating the unique atom that is true there. So, for instance, p would indicate that agent* 1 *owns the resource. Let us identify the state names with their associated atoms, and stipulate that $S^0 = \{p\}$: initially, agent 1 owns the resource. He can pass it on to 2 (leading us to the state where q is true) or to 3 (making r true). Agents 2 and 3 can also decide to keep the resource for themselves. We consider a norm η depicted by the dotted arrows (recall that a norm represents the "forbidden" transitions). This law is supposed to enforce that every agent will eventually get the resource. Let us identify a number of possible objectives associated with η:*

1. $\varphi_1 = \mathsf{A}\Diamond r$
   On every path, 3 *will eventually own the resource;*
2. $\varphi_2 = \mathsf{E}\Box\mathsf{E}\Diamond r$
   On some path, it is always the case that on some continuation r will eventually hold.
3. $\varphi_3 = \bigwedge_{x=p,q,r} \mathsf{A}\Box(x \rightarrow \mathsf{A}\Diamond\neg x)$
   Nobody keeps the resource forever.
4. $\varphi_4 = \mathsf{A}\Diamond(p \wedge (\mathsf{A}\Diamond(q \wedge \mathsf{A}\Diamond r)))$
   On every path, 1 will eventually obtain the resource, after which 2 will eventually obtain it, after which finally 3 will obtain it.
5. $\varphi_5 = \mathsf{A}\Diamond(p \wedge (\mathsf{A}\Diamond(q \wedge \mathsf{A}\Diamond(r \wedge \mathsf{A}\Diamond p))))$
   As φ_4, but back to 1 again.
6. *Consider the following fairness property $f(1)$ for agent* 1*:*

   $$p \rightarrow \mathsf{A}\bigcirc(\neg p \wedge \mathsf{A}\bigcirc\neg p)$$

   In words: if agent 1 has the resource, he will not have it in the next two rounds. Define $f(3)$ similarly with respect to r. Consider finally $\varphi_6 = (\mathsf{A}\Box f(1)) \vee (\mathsf{A}\Box f(3))$: computations are either fair with respect to 1 or to 3.

This gives us the following.

1. *We have the following Banzhaf scores: $\sigma_1 = 0, \sigma_2 = 4$ and $\sigma_3 = 0$. Note that for this objective, $swing(C, 2) = 1$ for every C with $2 \notin C$: first of all, agent 2 is* needed *to fulfill φ_1 (if 2 does not abide to η, he can keep the resource forever) and also sufficient: if* 2 *does not hang on to the resource or give it back to* 1*, agent* 3 *will eventually get it. Hence, we have $swing(C, i) = 0$ for $i = 1, 3$: since agent 2 is necessary and sufficient to make φ_1 true, agents 1 and 3 will never be in a swing position. Since $2^{n-1} = 4$ in this example, we have $\mu_1 = \mu_3 = 0$ and $\mu_2 = 1$. The Banzhaf indices β_i here equal the Banzhaf measures μ_i. Finally, the Shapley-Shubik indices are $\varsigma_1 = \varsigma_3 = 0$, and $\varsigma_2 = \frac{6}{6} = 1$.*
2. *This is an extreme case: note that φ_2 is true both in K and in $K \dagger \eta$, in other words, it does not matter who keeps to the norm and who does not. Consequently, $\sigma_i = \mu_i = \beta_i = \varsigma_i = 0$ for all $i \in A$.*
3. *This objective will be true iff both* 2 *and* 3 *comply with the norm. So, $swing(C, 2) = 1$ iff $3 \in C$ and $2 \notin C$, which yields $\sigma_2 = 2$. Similarly, $\sigma_3 = 2$, and obviously we have $\sigma_1 = 0$. This straightforwardly gives $\mu_1 = 0$ and $\mu_2 = \mu_3 = \frac{1}{2}$, $\beta_1 = 0$ and $\beta_2 = \beta_3 = \frac{1}{2}$. Also, $\varsigma_1 = 0$ and $\varsigma_2 = \varsigma_3 = \frac{3}{6} = \frac{1}{2}$.*
4. *Compliance to the norm by 1 and 2 is necessary and sufficient. Thus, the situation is similar to φ_3, and thus $\sigma_1 = \sigma_2 = 2$ and $\sigma_3 = 0$; $\mu_1 = \mu_2 = \frac{1}{2}$ and $\mu_3 = 0$; and $\varsigma_1 = \varsigma_2 = \frac{1}{2}$ and $\varsigma_3 = 0$.*
5. *Compliance to the norm by all agents is sufficient and necessary for φ_5. Hence, for every i, we have $swing(A \setminus \{i\}, i) = 1$, and hence $\sigma_i = 1$, $\mu_i = \frac{1}{4}$, and $\beta_i = \frac{1}{3}$. It is also not hard to see that $\varsigma_i = \frac{2}{6} = \frac{1}{3}$ for every $i \in A$.*
6. *This example illustrates that the Banzhaf index can be different from the Shapley-Shubik index as measures of power in normative systems. Observe that we need either* agents 1 and 2*, or* agents 1 and 3*, to comply to the norm if the objective φ_6 is to hold. Hence, we have that $swing(C, 1) = 1$ iff $C \in \{\{2\}, \{3\}, \{2, 3\}\}$ and $swing(C, 2) = 1$ iff $C = \{1\}$ and, finally, $swing(C, 3) = 1$ iff $C = \{1\}$. This gives $\sigma_1 = 3, \sigma_2 = 1 = \sigma_3$. For the*

Banzhaf measure this gives $\mu_1 = \frac{3}{4}$ *and* $\mu_2 = \mu_3 = \frac{1}{4}$*, and the Banzhaf index is* $\beta_1 = \frac{3}{5}$ *and* $\beta_2 = \frac{1}{5} = \beta_3$*. For the Shapley-Shubik index, we find* $\varsigma_1 = \frac{4}{6}$ *and* $\varsigma_2 = \frac{1}{6} = \varsigma_3$*. Note that* $\beta_1 = \frac{18}{30} < \frac{20}{30} = \varsigma_1$*. (In* ς_1*, 1 "gets a point twice" when considering 2 and 3: i.e. for 231 and 321, while in* β_1*, he only collects "one point" for joining* $\{2, 3\}$*.)*

An obvious question is whether structural properties of social systems (restricted forms of Kripke structure or objective formula) yield any information about power measures. First, we have the following.

Theorem 2. *Let* $S = \langle K, \varphi, \eta \rangle$ *be a social system such that* $\varphi \in L^u$*,* $K \not\models \varphi$*, and* $\eta \in \mathit{eff}(K, \varphi)$*. Then there is a player with a positive Banzhaf score.*

Next, let us consider how the Banzhaf score in particular is related to the logical structure of an objective formula. We let $S_j = \langle K, \varphi_j, \eta \rangle$ for $j = 1, 2, 3$ be social systems with K and η identical for each S_j. We will write $\nu_j(C)$ for $\nu_{S_j}(C)$, and σ_i^j for σ_i in game $G(S_j)$ (similarly for swing^j). If K and η are clear from the context, we will sometimes write $C \models \varphi$ for $K \dagger (\eta \upharpoonright C) \models \varphi$ and $C + i \models \varphi$ for $K \dagger (\eta \upharpoonright (C \cup \{i\})) \models \varphi$.

Theorem 3.

1. *if* $\varphi_1 \in \{\top, \bot\}$*, then no player is a swing player for any coalition, and hence* $\sigma_i = 0$ *for all i. The same is true for* φ_1 *being an objective formula, i.e., a Boolean combination of atoms from* Φ*: nobody can change the current state of affairs.*

2. *Suppose* $\varphi_1 = \neg\varphi_2$*. Note that, although we have* $\nu_1(C) = 1 \Rightarrow \nu_2(C) = 0$*, the other direction only holds if there is one unique starting state* $s^0 \in S^0$*, if there is an initial state in* $K \dagger (\eta \upharpoonright C)$ *where* φ_1 *is true and one where it is false, then we have* $\nu_1(C) = \nu_2(C) = 0$*.*

3. $\varphi_3 = \varphi_1 \wedge \varphi_2$*. We have* $\mathit{swing}^1(C, i) = 1 = \mathit{swing}^2(C, i) \Rightarrow \mathit{swing}^3(C, i) = 1$*, but it is of course possible that* $\sigma_i^3 = 0$ *while* $\sigma_i^1, \sigma_i^2 > 0$*, and also that* $\sigma_i^3 > 0$ *while* $\sigma_i^1 = \sigma_i^2 = 0$*.*

4. $\varphi_3 = \varphi_1 \vee \varphi_2$*. Again, when* $\mathit{swing}^1(C, i) = 1 = \mathit{swing}^2(C, i)$ *then* $\mathit{swing}^3(C, i) = 1$*, but the other way around does not hold. In case that* S^0 *is a singleton, then* $\mathit{swing}^3(C, i) = 1 \Rightarrow \mathit{swing}^2(C, i) = 1$ *or* $\mathit{swing}^1(C, i) = 1$*. Also for this singleton-case, let* $z = \mid \{C \subseteq A \mid C \not\models \varphi_1 \vee \varphi_2 \ \& \ C + i \models \varphi_1 \wedge \varphi_2\} \mid$*. Then* $\sigma_i^3 = \sigma_i^1 + \sigma_i^2 - z$*.*

5. $\varphi_3 = \mathsf{E}\bigcirc\varphi_2$*. This is an interesting case: if* $\sigma_i^3 > 0$*, it means that there is a coalition C that cannot enforce a path with a certain property, but when in addition i refrains from doing certain actions, such a path becomes available! If the reader doubts that this is an actual possibility in our framework, we offer* $\varphi_2 = \mathsf{A}(\mathit{alive}\,\mathcal{U}\,\mathit{old})$ *as an example. It is easy to construct a model and normative system where* $C \not\models \varphi_3$ *but* $C + i \models \varphi_3$[1]*. A tongue in cheek interpretation: if i does not refrain from smoking, there is no path in which he is guaranteed to get older than 65, while, if he would give up his habit, there is a path where he would certainly live to be old.*

4.1 Complexity of Power Indices

Now that we have some idea of how the measures described above may be applied in multi-agent systems, it is both natural and important to consider computational issues. Our first result is as follows:

Theorem 4. *Given a social system* $S = \langle K, \varphi, \eta \rangle$ *and agent i in K, computing the Banzhaf score* σ_i *for i in the corresponding coalitional game* $G(S)$ *is #P-complete.*

Proof. For membership of #P, consider a non-deterministic Turing machine that guesses a coalition $C \subseteq A \setminus \{i\}$, and accepts iff both $K \dagger (\eta \upharpoonright C) \not\models \varphi$ and $K \dagger (\eta \upharpoonright (C \cup \{i\})) \models \varphi$. Hence the number of computations on which this machine accepts will be the number of coalitions for which i is a swing, i.e., the Banzhaf score σ_i.

We now prove that computing the Banzhaf score is #P-hard, by a reduction from #SAT [10, p.169], the problem of counting the satisfying assignments of a given Boolean formula Ψ. Let $x_1, \ldots, x_k$ be the Boolean variables of Ψ. The reduction is as follows. For each Boolean variable x_i we create an agent a_i, and in addition we create two further agents, d and e. We also create Boolean variables corresponding to the variables $x_1, \ldots, x_k$ of the input instance Ψ, and two additional variables, y and z. We create $3k + 5$ states, and create the transition relation R and associated agent labelling α and valuation V as illustrated in Figure 2: inside states are the propositions true in that state, while arcs between states are labelled with the agent associated with the transition. Let $S^0 = \{s_1\}$ be the singleton initial state set. We have thus defined the Kripke structure K. For the remaining components, define $\eta = \{(s_1, s_3), (s_4, s_6), \ldots, (s_{3k+1}, s_{3k+3}), (y, y)\}$. Let Ψ^* be the formula obtained from Ψ by systematically replacing each Boolean variable x_i by $(\mathsf{E}\Diamond x_i)$. We transform the propositional input formula Ψ into a CTL formula χ representing an objective as follows:

$$\chi \mathrel{\hat{=}} \Psi^* \wedge \mathsf{A}\Box(y \rightarrow \mathsf{A}\bigcirc\mathsf{A}\Box z).$$

In words: 'Ψ^* holds and whenever y holds, the system continues to be always in z'. Finally, set the agent whose Banzhaf score is to be computed to e. Now, consider a coalition C such that $\mathit{swing}(C, e)$: we claim that C defines a satisfying assignment for Ψ. First, since the first conjunct in the definition of χ is satisfied, then C must correspond to a satisfying assignment for Ψ (the second conjunct in the definition of χ can only be achieved with the compliance of e). Conversely, given a satisfying assignment for Ψ, let C denote the corresponding coalition in the social system defined by the reduction. Then e will be a swing player for C; to see this, the compliance of C will ensure that the first conjunct in the definition of χ will be satisfied. However, the compliance of e is required to ensure that the second conjunct is satisfied. Thus, computing the Banzhaf score in the setting given is #P-hard. □

We will say a problem is *#P-equivalent* if it is #P-complete with respect to Turing reductions. We have:

Theorem 5. *Given a social system* $S = \langle K, \varphi, \eta \rangle$ *and agent i in K, the following problems are #P-equivalent: computing the Banzhaf index* β_i*; computing the Banzhaf measure* μ_i*; and computing the Shapley-Shubik index* ς_i*.*

[1] Take, e.g., four states $\{s_0, t, u, v\}$, transitions $(s_0, t), (t, u), (t, v)$, let *alive* be true only in t and *old* only true in v, let $\eta = \{t, u\}$, $A(t, u) = i$, and let $C = \emptyset$.

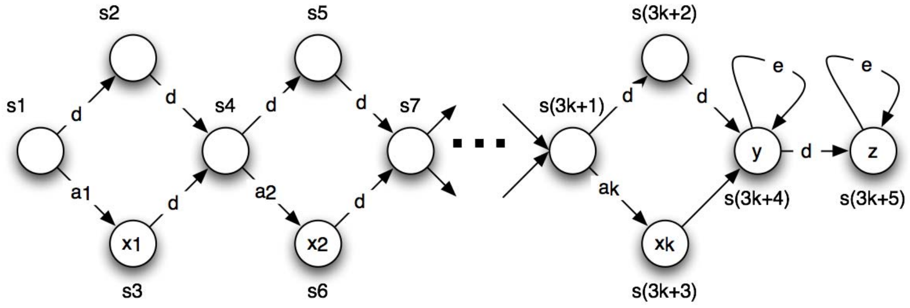

Figure 2: Illustrating the reduction for Theorem 4.

Proof. We show the proof for μ_i; the cases for β_i and ς_i are variations of essentially the same argument. For #P-easiness, note that computing the Banzhaf score of i, (i.e., the numerator in equation (2)), is in #P. It follows that computing μ_i in the setting given is #P-easy, since this value can be computed in polynomial time by a deterministic Turing machine with access to a #P oracle, i.e., one for computing σ_i. Now, if we had an efficient method for computing the Banzhaf measure, we would have an efficient method for computing the Banzhaf score (multiply μ_i by 2^{n-1}); but by Theorem 4 this problem is #P-hard. It follows that computing μ_i is #P-hard. □

We say that a player i is a *dictator* in a social system if $\mu_i = 1$, and a *dummy* if $\mu_i = 0$. In other words, a player is a dictator if its compliance with the normative system is both necessary and sufficient for the normative system to achieve the objective. If a player is a dummy, then its compliance, or otherwise, has no effect on whether the objective is achieved: the compliance of a dummy never makes any difference.

Theorem 6. *Given a social system $S = \langle K, \varphi, \eta \rangle$ and agent i in K, the following problems are co-*NP*-complete: checking whether $\sigma_i = 0$; checking whether $\mu_i = 0$; checking whether $\mu_i = 1$; checking whether $\beta_i = 0$; checking whether $\beta_i = 1$; checking whether $\varsigma_i = 0$; and checking whether $\varsigma_i = 1$.*

Proof. We show the proof for checking that $\sigma_i = 0$; the other cases are similar arguments and constructions. Consider the complement problem, i.e., the problem of checking whether $\sigma_i > 0$. Membership of NP is obvious: guess a coalition $C \subseteq A \setminus \{i\}$ and verify that i is a swing player for C. For hardness, we can reduce SAT, using the construction of Theorem 4: check whether e is a swing player for some coalition. □

Finally, it is interesting to consider the problem of *relative* power: given two agents $i, j \in A$ and a power index $M \in \{\sigma, \mu, \beta, \varsigma\}$, we write $i \succ_M j$ to mean $M_i > M_j$.

Theorem 7. *Given a social system $S = \langle K, \varphi, \eta \rangle$, agents i, j in K, and power measure $M \in \{\sigma, \mu, \beta, \varsigma\}$, it is* NP*-hard to decide whether $i \succ_M j$.*

Proof. Consider the case for $M = \sigma$. We can reduce from the problem of checking whether $\sigma_i > 0$, i.e., the complement of the problem we consider in Theorem 6, which from Theorem 6 is NP-complete. We simply add a new, "redundant", agent j so that $\sigma_j = 0$, giving $i \succ_\sigma j$ iff $\sigma_i > 0$. To define j, we simply associate no edges in the transition relation with j, so that j's compliance (or otherwise) to any normative system never makes any difference to the success or failure of any objective. The cases for $M \in \{\mu, \beta, \varsigma\}$ are similar. □

4.2 Tractable Instances

Interpreted according to the standard conventions of computational complexity, the results we presented above are negative: they indicate that computing power indices for normative systems in general is computationally complex. It is therefore obvious to ask whether there are any natural cases where computing these power indices becomes easy (polynomial time computable).

Minimality

The first case we consider concerns *minimal* normative systems (cf. [9]). We say that a social system $S = \langle K, \varphi, \eta \rangle$ is minimal if $K \dagger \eta \models \varphi$ but there does not exist an $\eta' \subset \eta$ such that $K \dagger \eta' \models \varphi$. In other words, in a minimal social system it is essential that *all* forbidden transitions remain unused: failing to observe any of the requirements in a minimal social system will result in the failure of the normative system. If $S = \langle K, \varphi, \eta \rangle$ is minimal we say that η is a *minimal norm* for K, φ. Now, given this, we can prove the following:

Theorem 8. *If $S = \langle K, \varphi, \eta \rangle$ is a minimal social system, then for each $i \in A(\eta)$, the values σ_i, μ_i, β_i, and ς_i are polynomial time computable. In fact, letting $m = |A \setminus A(\eta)|$, we have:*

$$\sigma_i = 2^m$$

from which μ_i and β_i may immediately be computed, and

$$\varsigma_i = \frac{1}{|A|!} \sum_{i=0}^{|A \setminus A(\eta)|-1} \binom{m-1}{i} (|A(\eta)| + i)!(m-1-i)!$$
$$= \frac{1}{|A(\eta)|}.$$

Proof. To see that $\sigma_i = 2^m$, from the fact that η is minimal, then agent $i \in A(\eta)$ is a swing player for coalition C iff $C \supset A(\eta) \setminus \{i\}$ and $i \notin C$. There are 2^m such coalitions. The case for ς_i follows from a similar argument, considering the number of possible permutations of A in which the agents $A(\eta) \setminus \{i\}$ all precede i (the numerator of the first expression for ς_i), which after simplification yields $\varsigma_i = \frac{1}{|A(\eta)|}$. □

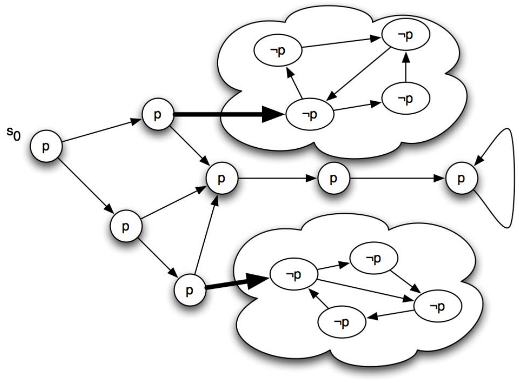

Figure 3: An example bridge normative system, in which the initial state is s_0, the objective to ensure that p is always true, and forbidden transitions are indicated by heavy arrows.

Of course, this easy computation is only feasible if ones knows that the social system is minimal, and in general, it is computationally hard to check this. If we do not know that the normative system in question is minimal, therefore, we cannot necessarily apply this idea. However, in some very natural cases, checking minimality can be straightforward. We here consider *bridge* and *tree* normative systems.

Bridge Normative Systems

The term "bridge" here derives from the way the term is used in graph theory [5, p.558]. The idea of a bridge normative system is as follows. Suppose we have an objective $\mathsf{A}\,\Box p$, and within the Kripke structure K, we have a bridge arc leading into a connected component (the "bad region") in which every state satisfies $\neg p$. Then this arc – the bridge – is an obvious candidate for a normative system to ensure $\mathsf{A}\,\Box p$: by forbidding this transition, we prevent the possibility of entering the bad region. The idea is illustrated in Figure 3.

Formally, if $S = \langle K, \mathsf{A}\,\Box\varphi, \eta\rangle$ is a social system (note the restricted form of objective), then we will say η is a *bridge* normative system if:

- for every arc $(s, s') \in \eta$, s is reachable from some member of S^0;
- for every arc $(s, s') \in \eta$, if we remove (s, s') from K then the component of R rooted at s' will be disconnected from every member of S^0 (i.e., (s, s') is a bridge);
- for every arc $(s, s') \in \eta$, then for every state s'' reachable in R from s', we have $K, s'' \models \neg\varphi$;
- for every arc $(s, s') \in \eta$, we have $K, s \models \varphi$.

Now, we have the following:

Theorem 9. *Given a social system $S = \langle K, \mathsf{A}\,\Box\varphi, \eta\rangle$ such that $\eta \in \mathit{eff}(K, \varphi)$, then if η is a bridge normative system, then η is minimal.*

Theorem 10. *Given a social system $S = \langle K, \mathsf{A}\,\Box\varphi, \eta\rangle$ such that $\eta \in \mathit{eff}(K, \varphi)$, checking whether η is a bridge normative system can be done in polynomial time.*

It follows that bridge normative systems represent a case where we can easily compute power indices.

Trees

We assume familiarity with the notion of a tree $T = \langle S, r, R\rangle$, with root r and domain S. A Kripke structure $K = \langle S, \{s_0\}, R, A, \alpha, V\rangle$ will be called a Kripke tree if $\langle S, \{s_0\}, R'\rangle$ is a tree, where $R' = R \setminus \{(s, s) : s \in L\}$ and $L = \{s : (s, t) \in R \Rightarrow t = s\}$. L is the set of *leaves* (note that R' is the same as R only with the self-loops at the leaves, a necessity due to the totality requirement, removed in order to get a proper tree). The nodes $S \setminus L$ which are not leaf nodes are called decision nodes. We will focus on one type of goal only: $\varphi = \mathsf{A}\Diamond g$ where g ("good and terminated") is a propositional formula. A final restriction on Kripke trees is that we assume that g is only true in (some of the) leaves, and is true in at least one leaf. In this section we are interested in social systems $S = \langle K, \mathsf{A}\Diamond g, \eta\rangle$ where K is a Kripke tree. An example of a Kripke tree is shown in Figure 4.

Theorem 11. *If K is a Kripke tree, then there is a unique minimal norm η_{min} for $K, \mathsf{A}\Diamond g$. Given K, η_{min} can be constructed in linear time.*

Proof. We describe an algorithm for constructing η_{min} as follows. The algorithm goes through the tree in two passes. The first pass is a depth-first traversal starting at the root, marking each node with one of $\{+, -, =\}$ in a post-order sequence as follows: for each leaf $\ell \in L$, mark ℓ with $+$ iff $\ell \models g$ and with $-$ otherwise. For each decision node s, mark s with $+$ if all its children are marked with $+$, with $-$ if all its children are marked with $-$, and with $=$ otherwise. (The marking is illustrated in Figure 4).

The second pass is also a depth-first traversal starting at the root, where for each node $s \in S$ a set η_s is defined in a post-order sequence as follows. $\eta_\ell = \emptyset$ when $\ell \in L$. For each $s \in S \setminus L$, let η_s be defined as follows from the markings and sets η_t of its child nodes t. For each child node t of s ($(s, t) \in R'$), if t is marked with $=$ let $\eta_t \subseteq \eta_s$; if t is marked with $-$ then let $(s, t) \in \eta_s$.

Finally, let $\eta_{min} = \eta_{s_0}$. It is easy to see that no leaf node where g is not safisfied is reachable from the root in $K \dagger \eta_{min}$, and thus that $K \dagger \eta_{min} \models \mathsf{A}\Diamond g$. Minimality follows from construction. Uniqueness follows from the totality requirement: if $(s, t) \in \eta_{min}$ then (by construction) none of the leaves of the subtree rooted at t satisfy g, and if a normative system did not include (s, t) then it would have to remove *every* outgoing transition from at least one node in that subtree. Note that each depth-first traversal is done in time proportional to $|S| + |R|$. □

It follows that for Kripke trees and objectives of the form $\mathsf{A}\Diamond g$, not only can minimality and therefore the power indices be computed in polynomial time (Theorem 8), but it is also the case that the (unique) minimal norm ensuring the goal can be *synthesised* in polynomial time.

Example 2. *Let K be the Kripke tree with $A = \{1, 2, 3, 4, 5\}$ illustrated in Figure 4. The minimal norm is*

$$n_{min} = \{(b, f), (c, i), (h, m)\}$$

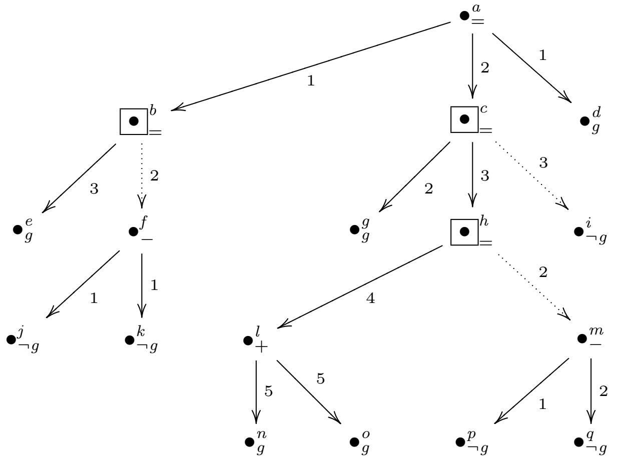

Figure 4: Example Kripke tree with states $\{a, \ldots, q\}$. Leaves are marked with the truth value of g. Decision nodes are marked with $\{+, -, =\}$ according to the algorithm in Theorem 11. The boxed decision nodes illustrate "critical" nodes.

(illustrated with dotted lines in the figure) and $A(\eta_{min}) = \{2, 3\}$. In the notation of Theorem 8, $m = 3$. Thus, for $j \in \{1, 4, 5\}$:

$$\begin{aligned}
&\sigma_2 = \sigma_3 = 8; \sigma_j = 0\\
&\mu_2 = \mu_3 = \tfrac{1}{2}; \mu_j = 0\\
&\beta_2 = \beta_3 = \tfrac{1}{2}; \beta_j = 0\\
&\varsigma_2 = \varsigma_3 = \tfrac{1}{5}; \varsigma_j = 0
\end{aligned}$$

5. DISCUSSION AND CONCLUSIONS

Our work builds on several very different areas: the use of CTL-like logics for reasoning about distributed and multi-agent systems [7], social laws for coordinating multi-agent systems [12, 13, 14, 1], cooperative game theory and power indices/voting theory [11, 8], and computational complexity [10]. To the best of our knowledge, the present paper is the first synthesis of these different domains; the closest work we know of is the seminal work of [3], who use power indices to analyse network flow games, with the goal of finding particularly important nodes or bottlenecks in the network. However, the work also seems related to research on the theory of influence [2]. The complexity of cooperative solution concepts such as the Shapley value was originally studied, for a number of coalitional game representations, in [6], although it has been known since at least 1979 that computing the Shapley-Shubik index for weighted voting games is #P-complete [10, p.280].

Several issues suggest themselves for future work. Most obviously, it will be important to try to identify further tractable instances of the problems considered, focusing, for example on restricted classes of Kripke structures and CTL objectives. In addition, it would seem worth investigating the complexity of the problems considered in this paper for more succinct representations of Kripke structures, such as those used by model checking systems: we might expect the typical complexity of computing power indices to be at least PSPACE-hard for such representations. And finally, of course, more experience with the use of these measures in practical settings would be valuable.

6. REFERENCES

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. Normative system games. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2007)*, Honolulu, Hawaii, 2007.

[2] N. I. Al-Najjar and R. Smorodinsky. Pivotal players and the characterization of influence. *Journal of Economic Theory*, 92(2):318–342, 2000.

[3] Y. Bachrach and J. S. Rosenschein. Computing the Banzhaf power index in network flow games. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2007)*, pages 335–341, Honolulu, Hawaii, 2007.

[4] J. F. Banzhaf III. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, 19(2):317–343, 1965.

[5] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms.* The MIT Press: Cambridge, MA, 1990.

[6] X. Deng and C. H. Papadimitriou. On the complexity of cooperative solution concepts. *Mathematics of Operations Research*, 19(2):257–266, 1994.

[7] E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science Volume B: Formal Models and Semantics*, pages 996–1072. Elsevier Science Publishers B.V.: Amsterdam, The Netherlands, 1990.

[8] D. S. Felsenthal and M. Machover. *The Measurement of Voting Power.* Edward Elgar: Cheltenham, UK, 1998.

[9] D. Fitoussi and M. Tennenholtz. Choosing social laws for multi-agent systems: Minimality and simplicity. *Artificial Intelligence*, 119(1-2):61–101, 2000.

[10] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman: New York, 1979.

[11] M. J. Osborne and A. Rubinstein. *A Course in Game Theory.* The MIT Press: Cambridge, MA, 1994.

[12] Y. Shoham and M. Tennenholtz. On the synthesis of useful social laws for artificial agent societies. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*, San Diego, CA, 1992.

[13] Y. Shoham and M. Tennenholtz. On social laws for artificial agent societies: Off-line design. In P. E. Agre and S. J. Rosenschein, editors, *Computational Theories of Interaction and Agency*, pages 597–618. The MIT Press: Cambridge, MA, 1996.

[14] W. van der Hoek, M. Roberts, and M. Wooldridge. Social laws in alternating time: Effectiveness, feasibility, and synthesis. *Synthese*, 156(1):1–19, May 2007.