

Game-Theoretic Aspects of Designing Hyperlink Structures

Nicole Immorlica, Kamal Jain, and Mohammad Mahdian

Microsoft Research, One Microsoft Way, Redmond, WA.
{nickle,kamalj,mahdian}@microsoft.com

Abstract. We study the problem of designing the hyperlink structure between the web pages of a web site in order to maximize the revenue generated from the traffic on the web site. We show this problem is equivalent to the well-studied setting of infinite horizon discounted Markov Decision Processes (MDPs). Thus existing results from that literature imply the existence of polynomial-time algorithms for finding the optimal hyperlink structure, as well as a linear program to describe the optimal structure. We use a similar linear program to address our problem (and, by extension all infinite horizon discounted MDPs) from the perspective of cooperative game theory: if each web page is controlled by an autonomous agent, is it possible to give the individuals and coalitions incentive to cooperate and build the optimal hyperlink design? We study this question in the settings of transferrable utility (TU) and non-transferrable utility (NTU) games. In the TU setting, we use linear programming duality to show that the core of the game is non-empty and that the optimal structure is in the core. For the NTU setting, we show that if we allow “mixed” strategies, the core of the game is non-empty, but there are examples that show that the core can be highly inefficient.

1 Introduction

As electronic commerce begins to dominate the business model of many companies, the design of an efficient and revenue-maximizing web site is of increasing importance. A major component of web site design is the selection of the hyperlink structure among the web pages. A web designer can be likened to a city planner, building hyperlink structure so as to steer traffic in a globally optimal manner. One consideration, which is of particular importance for web sites whose objective is to provide information for the users, is to facilitate the navigation through the contents of the web site. The other consideration, in particular for designing e-commerce web sites, is to present links on each page in order to direct a surfer through a path of high revenue. The latter objective is the focus of this paper.¹

We provide a graph-theoretic model for this problem. Web pages generate varying amounts of revenue, perhaps through advertisements or product sales.

¹ As we will see in Section 5, our framework can be generalized to accommodate content-related constraints as well.

Additionally, web pages display hyperlinks to some other pages on the web site. Each possible hyperlink has a transition probability representing the probability that a surfer clicks on the hyperlink conditional on the other links on the page. The web designer now must select a subgraph which maximizes the expected revenue of a random walk. As we will show, the stated problem is in fact equivalent to infinite horizon discounted Markov Decision Processes (MDPs) (see [7]). Thus, the value iteration algorithm for MDPs [7] can be used to compute the optimal hyperlink structure efficiently.

In this paper, we use a linear-programming formulation for MDPs to give us insight into some game-theoretic aspects of the web design question (and of MDPs in general). Often, in large companies like Amazon or MSN, web pages are controlled by distinct (and sometimes even competing) profit centers, each responsible for their own profit and loss (P&L) account. It is not reasonable to assume that a particular profit center, or group of profit centers, will comply with the optimal web design at its own expense. Rather, it is necessary to divide the total revenue of the web site among the profit centers to ensure stability. We formulate our concern as a transferrable utility game and use insights from cooperative game theory and the LP formulation of the problem to compute an allocation scheme in the core of the game. This implies that there is always a way to divide revenue among profit centers such that the optimal web site design is stable, i.e. each group of profit centers receives a total revenue at least as large as the revenue they would be able to extract if they deviate as a coalition. We further study the non-transferrable utility game which is more suitable for situations where monetary transfer between agents managing different web pages is not possible, i.e., when each web page receives precisely the revenue it generates. We prove that in this case, if “mixed” strategies are allowed, the core is non-empty, i.e. there is a web site design where no profit center (or group of profit centers) can deviate and increase its revenue. However, the efficient web site design need not be in the core of the game, and furthermore, we show that there are examples where the revenue of the core is worse than the optimal solution by an arbitrary factor.

Our work bears some similarity to the long-standing tradition of network formation games in the economics literature (see [5] for a survey). This literature takes the standpoint that social networks play a key role in many economic settings, including labor markets [4], international free trade agreements [3], and peering and transit relations on the Internet [1]. As such, much effort has been invested in understanding economic incentives facing agents forming links in these social networks. A variety of value functions have been proposed to describe the effect of particular network structures on individuals, and our framework can be adopted to study these settings as well. However, for many of them, the computational questions remain open.

The rest of this paper is organized as follows. In Section 2, we define the model and its relationship to MDPs. In Section 3, we present a linear-programming formulation for describing the optimum (revenue-maximizing) web site design. In Section 4, we use the LP presented in Section 3 to discuss game-theoretic

problems and prove that the cores of both the transferrable utility and non-transferrable utility games are non-empty. We conclude with the discussion of a multitude of interesting generalizations and open questions in Section 5.

2 Model

We model a web site as a directed graph $G = (N, E)$. Each node $i \in N$ is a web page. We denote the number of nodes by $n = |N|$. An edge ij exists from node i to node j if page i links to page j . We assume that this graph contains no self-loop, i.e., a web page does not link to itself.

A web surfer is represented by a random walk on this graph. For each page j , there is a probability p_j that the surfer starts surfing from page j . For each page i , set $S \subset N \setminus \{i\}$ of other pages, and page $j \in S$, there is a probability $p_{ij,S}$ that a surfer on page i follows a hyperlink to page j , assuming that the set of pages linked from page i is S . We assume that for all i and $S \subset N \setminus \{i\}$, $\sum_{j \in S} p_{ij,S} \leq 1 - \delta$ for some positive constant $\delta > 0$, i.e., in each step there is a non-zero probability that the surfer exits the web site.

We define a revenue for a random walk on the web site. The simplest way to do this is to assign a revenue r_j to each page j (this would correspond to the expected revenue that a surfer visiting page j would generate for the web site owner, perhaps from the advertisement on the page or by buying a product on the page), and define the expected revenue of a random walk as the sum, over all j , of r_j times the expected number of times that the random walk visits j . In this paper, we consider a more general model where the revenues are assigned to edges instead of vertices: for each hyperlink ij , there a value $r_{ij,S}$ representing the expected revenue generated for page j by a web surfer who has followed link ij when the links on page i were S . The total revenue is defined as the sum, over all edges ij in the graph, of $r_{ij,S}$ times the expected number of times the random walk traverses the edge ij . Notice that this is a strictly stronger model, since setting $r_{ij,S} = r_j$ for all i and S would be equivalent to assigning revenues to vertices (of course, we also need to add the value $\sum_j p_j r_j$ for the revenue of the first page the surfer visits). Assigning revenues to edges enables us to model situations where the conversion rate of a user depends on the web page she is coming from, and will be useful in modelling content-related constraints (as discussed in Section 5). Note that we defined the total revenue by multiplying $r_{ij,S}$'s by the expected *number of times* the random walk takes the corresponding edge, as opposed to the probability that the random walk takes this edge. This means that if the random walk visits a vertex twice, it will benefit the web site owner twice. This is a realistic assumption in many situations, e.g., where the revenue is generated from “per-impression” advertisements. For a discussion of alternative models, see Section 5.

2.1 The Case of No Externalities

The above model is strong enough to model situations where the probability that a surfer clicks on a link to page j placed on page i depends not only on i

and j , but also on the set of other links on the page i . In economic terminology, this means that we can model *externalities* among the links placed on a page i . An interesting and important special case is the case of *no externalities*. In this case, each page has limited real-estate in which it can display links, and so each node i can have out-degree at most k_i (a parameter). For each $i, j \in N$, there is a probability p_{ij} that a surfer on page i follows a hyperlink to page j , if such a link exists. We assume that for all i , and for any set S of k_i pages, the sum $\sum_{j \in S} p_{ij} \leq 1 - \delta$, so these probabilities define a random walk with exit probability at least δ in each step.² See Section 5 for a discussion of other models where instead of (or in addition to) the limit k_i on the number of links, there is a cost associated with placing each link.

2.2 Equivalence to Markov Decision Processes

A Markov Decision Process (MDP) is a common construct used to describe scenarios with sequential decision-making processes. An MDP consists of a set of states \mathcal{S} , a set of actions $\mathcal{A}(s)$ for each state $s \in \mathcal{S}$, and a revenue $r_{a,s}$ for each action/state pair.

In each iteration of an MDP, the system is in a state s , and an action $a \in \mathcal{A}(s)$ must be chosen. Actions induce a probability distribution over future states, and revenue $r_{a,s'}$ is interpreted as the revenue of taking action a given that the resulting state is s' . The goal is to choose an action for each state which maximizes the total (expected) revenue of the system over time. In infinite horizon discounted MDPs, the total revenue is calculated with respect to a discounting factor (λ), i.e. the (expected) revenue r in the t 'th iteration contributes $\lambda^t r$ to the total (expected) revenue.

That the model introduced above is equivalent to infinite horizon discounted MDPs can be seen by equating the set of states \mathcal{S} with the web pages N . The actions $\mathcal{A}(i)$ for a web page $i \in N$ are subsets S of other pages. By adding a “terminal” web page and links from each page to the terminal page with appropriate probabilities, we can ensure that the sum of the probabilities of the links leaving each page is precisely $1 - \delta$. Given this assumption, the induced probability distribution for taking the action S at state i of the MDP can be defined as $p_{ij,S}/(1 - \delta)$. For action $S \in \mathcal{A}(i)$, a revenue of $r_{ij,S}$ is generated given that the action resulted in future state j . The discounting factor λ is equal to $1 - \delta$.

Due to the above equivalence, one can easily adapt known algorithms for MDPs, such as the value iteration algorithm, to compute the optimal hyperlink design efficiently. Furthermore, it can be easily seen that all of the results of this paper can be applied to general infinite horizon discounted MDPs.

² Strictly speaking, in this model there is still externality among the links, since placing each link further limits the number of other links that can be placed on the page. However, this is the only form of externality allowed in this case.

3 Linear Programming Formulation

In this section, we present a linear program which describes the revenue-maximizing hyperlink structure. For simplicity of presentation, we describe the program in the case of no externalities.

The optimization question facing a web designer in our setting is to find a subgraph of the complete graph in which each node has degree at most k_i and the total revenue is maximized. This can be formulated as a mathematical program as follows. Let x_i be a variable representing the expected number of times a web surfer encounters node i and y_{ij} be an indicator variable for the existence of hyperlink ij . Thus, the expected number of times a web surfer traverses link ij is simply $x_i p_{ij} y_{ij}$. Relaxing the integrality constraint on y_{ij} , the problem then becomes

$$\max \sum_{i,j \in N} r_{ij} \cdot (x_i p_{ij} y_{ij}) \quad (1)$$

$$\text{s.t. } \forall j \in N : x_j \leq p_j + \sum_{i \in N} x_i p_{ij} y_{ij} \quad (2)$$

$$\forall i \in N : \sum_{j \in N} y_{ij} \leq k_i \quad (3)$$

$$\forall i, j \in N : 0 \leq y_{ij} \leq 1$$

$$\forall i \in N : x_i \geq 0.$$

Constraint 2 encodes the ‘‘conservation of flow’’: the expected number of times x_j a surfer visits node j can not be more than the expected number of times p_j he starts surfing from j plus the expected number of times $\sum_{i \in N} x_i p_{ij} y_{ij}$ that he enters j from a neighboring node. Constraint 3 encodes the out-degree constraint on a node i .

This mathematical program can be transformed to a linear program by performing the change of variables $z_{ij} = x_i y_{ij}$. This gives us the program

$$\max \sum_{i,j \in N} r_{ij} p_{ij} z_{ij} \quad (4)$$

$$\text{s.t. } \forall j \in N : x_j \leq p_j + \sum_{i \in N} p_{ij} z_{ij}$$

$$\forall i \in N : \sum_{j \in N} z_{ij} \leq k_i x_i$$

$$\forall i, j \in N : z_{ij} \leq x_i$$

$$\forall i \in N : x_i \geq 0$$

$$\forall i, j \in N : z_{ij} \geq 0$$

which is linear in the variables x_i and z_{ij} . In the next section, we show how to round an optimal fractional solution (x_i, z_{ij}) to LP 4 to a solution in which $z_{ij}/x_i \in \{0, 1\}$ for all $i, j \in N$. This shows that the above LP formulation exactly captures the hyperlink design problem, a fact that will be used in the next section to derive the game-theoretic results.

3.1 Rounding Technique

Consider an optimal fractional solution to LP 4. For all $i \in N$ such that $x_i > 0$ and all $j \in N$, define $y_{ij} = z_{ij}/x_i$. Notice if $y_{ij} \in \{0, 1\}$ for all $i, j \in N$, then we can use these y_{ij} to define a feasible hyperlink structure with optimal revenue.

Otherwise, let $G = (N, E)$ be the graph where edge ij exists if $y_{ij} > 0$ and has transitional probability $p_{ij}y_{ij}$. Consider an arbitrary node $i_0 \in N$ with at least one fractional out-going edge, i.e. for at least one j , $0 < y_{i_0j} < 1$. We “fix” this node without sacrificing any of the total revenue.

Lemma 1. *There is a graph G' with total expected revenue equal to G in which i_0 has exactly k_{i_0} integral out-links.*

Proof. In order to prove this claim, we will write the fractional out-links of i_0 in G as a convex combination of feasible integral out-links and show that one of these corresponding graphs has revenue at least that of G .

As G is an optimal fractional graph, we may assume that $\sum_j y_{i_0j} = k_{i_0}$. Thus, the $\{y_{i_0j}\}$ lie in the integral polytope described by $\sum_j y_{i_0j} = k_{i_0}$ and $0 \leq y_{i_0j} \leq 1$. Let $F_l \in \{0, 1\}^{|N|}$ be the vertices of this polytope, and note that each F_l has exactly k_{i_0} non-zero coordinates. We represent the $\{y_{i_0j}\}$ as a convex combination of these vertices $\sum_l \lambda_l F_l$ where $\sum_l \lambda_l = 1$ and $\lambda \geq 0$.

Consider the graph $G_l = (N, E_l)$ where i_0 only has links in F_l . In other words, $E_l = E - \{y_{i_0j}\} + \{i_0j : F_l(j) = 1\}$. Let R'_l be the expected revenue that a random walk in G_l starting at i_0 collects before returning to i_0 . Furthermore, let p_l be the probability that a random walk in G_l starting at i_0 returns to i_0 . Note $p_l < 1$ as there is an exit probability at each node. Thus, the total expected revenue R_l of a random walk starting from i_0 in G_l can be written as $R_l = R'_l + p_l R_l$, and so

$$R_l = \frac{R'_l}{1 - p_l}.$$

We would like to prove that for some l , the revenue R_l of G_l starting at i_0 is at least the revenue of G starting at i_0 . We can write the revenue R of G starting at i_0 in terms of R'_l as follows: by linearity of expectation, the expected revenue that a random walk in G starting at i_0 collects before returning to i_0 is simply $\sum_l \lambda_l R'_l$. Also, the probability of returning to i_0 is $\sum_l \lambda_l p_l$. Therefore, $R = \sum_l \lambda_l R'_l + \sum_l \lambda_l p_l R$, and so

$$R = \frac{\sum_l \lambda_l R'_l}{1 - \sum_l \lambda_l p_l}.$$

Using the fact that $\sum_l \lambda_l = 1$, we can rewrite R as

$$R = \frac{\sum_l \lambda_l R'_l}{\sum_l \lambda_l (1 - p_l)}$$

where we restrict the summation to the vertices F_l such that $\lambda_l > 0$. Using the fact that $(\sum_l a_l)/(\sum_l b_l) \leq \max_l (a_l/b_l)$ for any two sequences of positive reals

$\{a_l\}$ and $\{b_l\}$, we see for some l , the revenue of G_l starting at i_0 is at least the revenue of G starting at i_0 . Note that the revenue of a random walk starting from a node $j \neq i_0$ is the same in G and G_l until it reaches i_0 as we only changed the out-going links of i_0 . Therefore, we can conclude that the total revenue of G_l is at least that of G .

We can now proceed to “fix” iteratively all nodes i with fractional out-links to get an integral graph G with optimal revenue.

3.2 General Externalities between Links

We remark that all the results of this section can be extended to the general case by using the following mathematical programming formulation. Let $y_{i,S}$ be an indicator variable for the event that page i chooses to link to pages in S . As before, x_i represents the expected number of times a surfer visits page i . By convention, we define $p_{i,j,S} = 0$ for $j \notin S$.

$$\begin{aligned}
& \max \sum_{i,j \in N, S \subseteq N} r_{i,j,S} \cdot (x_i p_{i,j,S} y_{i,S}) & (5) \\
& \text{s.t. } \forall j \in N : x_j \leq p_j + \sum_{i \in N, S \subseteq N} x_i p_{i,j,S} y_{i,S} \\
& \forall i \in N : \sum_{j \in N, S \subseteq N} y_{i,S} \leq 1 \\
& \forall i, j \in N : 0 \leq y_{i,S} \leq 1 \\
& \forall i \in N : x_i \geq 0.
\end{aligned}$$

4 The Cooperative Hyperlink Design Game

Cooperative game theory, defined by von Neumann and Morgenstern in 1944 [10], studies games in which the primitives are actions taken by coalitions of players (see [6] for background on cooperative game theory). The setting defined in Section 2 can be interpreted as a cooperative game where the nodes of the graph (i.e., the web pages) are the players. Thus, each web page is owned by a individual self-motivated agent such as a profit center within a company. This individual seeks hyperlinks that maximize his own revenue, but may cooperate with other web page owners in doing so and thereby capitalize on the induced externalities between web pages. For simplicity of presentation, we again describe our results in the case of no externalities between links, although all our results extend easily to the general case using program 5.

We consider both a transferrable and non-transferrable utility setting. In a transferrable utility setting, the value generated by a coalition may be distributed in an arbitrary manner among the members of the coalition whereas in our non-transferrable setting, each node in a coalition receives only the revenue it generates³.

³ The formal definition of non-transferrable games allows for more general payoff vectors.

4.1 Cooperative Game with Transferrable Utility

In a transferrable utility game, the underlying assumption is that the revenue generated by a coalition may be shared among its members in any manner. A *transferrable utility* (TU) game is defined by a value function v which assigns to every possible coalition of players the value they can achieve. In our setting, the value $v(S)$ of a subset S of nodes is the value of the corresponding linear program 4 with variables restricted to the set S (i.e., the LP applied to the subgraph induced by the nodes in S). A solution of the game is a set of payoffs ξ_i , one for each player, such that $\sum_{i \in N} \xi_i = v(N)$.

We would like to define a notion which describes the stable solutions of the game. A standard such notion is that of the *core*, defined by Gillies, Shapley, and Shubik in a series of papers in the 1950s and 1960s. A solution is in the *core* of a coalitional game with transferrable utility if for all coalitions S , $\sum_{i \in S} \xi_i \geq v(S)$. Thus, the core is described by a set of linear inequalities.

Definition 1. *A set of payoffs ξ_i is in the core if $\sum_{i \in N} \xi_i = v(N)$ and for all $S \subset N$, $\sum_{i \in S} \xi_i \geq v(S)$.*

We prove that our game has a non-empty core. This claim can be proved using a famous theorem of Bondareva [2] and Shapley [9] which characterizes the games with non-empty cores. However, we provide a proof based on LP-duality to establish our algorithmic result for computing a solution in the core.

In order to write the dual of linear program 4, we assign variables α_j , β_i , and γ_{ij} corresponding to the first, second, and third inequality, respectively. The dual is then

$$\begin{aligned}
 \min \quad & \sum_{i \in N} \alpha_i p_i & (6) \\
 \text{s.t.} \quad & \forall j \in N : \alpha_j - k_j \beta_j - \sum_{i \in N} \gamma_{ij} \geq 0 \\
 & \forall i, j \in N : -\alpha_j p_{ij} + \beta_i + \gamma_{ij} \geq r_{ij} p_{ij} \\
 & \forall j \in N : \alpha_j \geq 0 \\
 & \forall i \in N : \beta_i \geq 0 \\
 & \forall i, j \in N : \gamma_{ij} \geq 0.
 \end{aligned}$$

We claim the payoffs $\xi_i = \alpha_i p_i$ are in the core. Clearly $\sum_{i \in N} \xi_i = \sum_{i \in N} \alpha_i p_i = v(N)$ by LP-duality. Also by LP-duality, to prove for all $S \subset N$, $\sum_{i \in S} \xi_i \geq v(S)$, we only need to show that the optimal solution $(\alpha_j, \beta_i, \gamma_{ij})$ to LP 6 is a feasible solution to LP 6 restricted to players in S . This follows easily as the inequalities of LP 6 restricted to the players in S are a subset of those in LP 6.

We have thus proved that our game has a non-empty core, and we can find a solution in this core in polynomial time.

4.2 Cooperative Game with Non-transferrable Utility

Transferrable utility games assume that the players are able to distribute the total utility in *any* manner. In many settings, such an assumption is unreasonable.

For example, in our setting, the performance of a profit center is often measured in terms of the amount of revenue it generates for the company, and there is no mechanism through which profit centers may share revenue prior to review. A non-transferrable utility game generalizes transferable utility games by studying situations such as these in which not all payoff vectors are feasible for a coalition.

A *non-transferrable utility* (NTU) game consists of a set N of players and for each coalition $S \subseteq N$ a set $\mathcal{V}(S) \subset \mathbb{R}^{|S|}$ of feasible payoff vectors for that coalition. The sets $\mathcal{V}(S)$ are assumed to satisfy some mild assumptions, namely: 1. $\mathcal{V}(S)$ is closed; 2. if $v \in \mathcal{V}(S)$, then for all $v' \in \mathbb{R}^{|S|}$ with $v' \leq v$ (coordinate-wise), $v' \in \mathcal{V}(S)$; and 3. the set of vectors in $\mathcal{V}(S)$ in which each player receives at least the utility he can achieve individually is a nonempty, bounded set. Intuitively, a solution to an NTU game with payoffs $v \in \mathcal{V}(N)$ is stable (in the core) if no coalition S can withdraw and achieve a payoff vector $v' \in \mathcal{V}(S)$ such that each member of S improves his payoff. For notational convenience, we will use $v|_S$ denote the vector in $\mathbb{R}^{|S|}$ whose coordinates are the coordinates of v restricted to the players in S . A vector $v \in \mathcal{V}(N)$ is in the *core* of the NTU game if there is no coalition S and vector $v' \in \mathcal{V}(S)$ such that $v' > v|_S$ (coordinate-wise). The following result of Scarf [8] states a condition under which an NTU game has a non-empty core.⁴ Let λ_S be a *fractional partition* λ_S of players, i.e., a set of coefficients $0 \leq \lambda_S \leq 1$ of subsets of N such that for all players i , $\sum_{S:i \in S} \lambda_S = 1$. An NTU game is called *balanced* if, for every fractional partition λ_S , a vector $v \in \mathbb{R}^{|N|}$ must be in $\mathcal{V}(N)$ if $v|_S \in \mathcal{V}(S)$ for all S with $\lambda_S > 0$.

Theorem 1. (*Scarf*) *A cooperative game with non-transferrable utility has a non-empty core if it is balanced.*

In our setting, the set $\mathcal{V}(S)$ consists of the payoff vectors v where v_i is (at most) the revenue of i in some hyperlink structure on S . More formally, $v \in \mathcal{V}(S)$ if and only if there is a (fractional) graph G on nodes S such that for each player $i \in S$, v_i is at most the expected revenue of i in G . Alternatively, we can state this condition using program 1: $v \in \mathcal{V}(S)$ if and only if there is a feasible solution (x_i, y_{ij}) to program 1 such that for each player $i \in S$, v_i is at most $\sum_j r_{ji} \cdot (x_j p_{ji} y_{ji})$ (the expected revenue of i). These sets $\mathcal{V}(S)$ clearly satisfy the assumptions stated above, and so our game is an NTU game. Here, we use Scarf's theorem to prove the following statement.

Theorem 2. *There is a fractional graph in the core of the web site game.*

Proof. Consider any fractional partition λ_S and payoff vectors $v(S) \in \mathcal{V}(S)$. Let v be the vector whose i 'th coordinate is the minimum over all S containing i of $v(S)|_i$. We prove that there is a fraction graph G whose corresponding payoff vector v' satisfies $v' \geq v$ (coordinate-wise) and thus the game is balanced. Our theorem then follows from Scarf's theorem.

Let $(x_{i,S}, z_{i,j,S})$ be a feasible solution to LP 4 on set S such that $v(S) \leq \sum_{j \in S} r_{ji} p_{ji} z_{ji,S}$. Consider the solution (x_i, z_{ij}) to LP 4 where $x_i = \sum_{S:i \in S} \lambda_S x_{i,S}$

⁴ This is a generalization of the result of Bondareva [2] and Shapley [9] which states a condition under which a TU game has a non-empty core.

and $z_{ij} = \sum_{S:i,j \in S} \lambda_S z_{ij,S}$. As (x_i, z_{ij}) is a convex combination of feasible solutions, it is also feasible and thus the corresponding graph G is a feasible fractional graph. Furthermore, the revenue v'_i of a node i in G is at least $\sum_{j \in N} r_{ji} p_{ji} z_{ji} = \sum_{S:i \in S} \lambda_S \sum_{j \in S} r_{ji} p_{ji} z_{ji,S} \geq \min_S v(S)|_i \geq v|_i$.

Fractional graphs can be thought of as the result of mixed strategies in link selection. In other words, if we allow a node i to have fractional out-links of total weight at most k_i (or probabilistically select k_i links according to their fractional weight), then the core is non-empty. See section 5 for a discussion regarding computational issues and “pure” strategies.

We end with a comment regarding the efficiency of the graphs in the core. Whereas the efficient (that is, revenue-maximizing) graph is in the TU core, this may not be the case for the NTU core. In fact, the solutions in the NTU core may be arbitrarily inefficient. As an example, consider the game on three nodes a , b , and c . Suppose $p_a = 1$ so a surfer always enters the site at node a . The revenue of any link entering node a is 1, b is 1, and c is R for an arbitrarily large R . Each node is allowed (fractionally) one out-link. The transition probabilities are $p_{ab} = 1/2$, $p_{ba} = 1/2$, $p_{ac} = 1/2$, and all other transition probabilities are 0. It is easy to check that the only solutions in the NTU core of this game include integrally the set of links $\{ab, ba\}$. However, the revenue of any such graph is constant while the efficient graph $\{ba, ac\}$ has arbitrarily large revenue.

5 Discussion

Our model and results are quite general and can be accommodated to handle a large number of scenarios. We discuss some of them here, and mention a few open questions.

Content-related restrictions. In many web sites, the link structure might be subject to certain content-related restrictions. For example, perhaps MSNBC is required to link to MSN search regardless of the transition probability. Our setting is general enough to handle a wide variety of such restrictions by appropriately setting certain $p_{ij,S}$ to zero. In the above example, if a link ij is required to appear, we can set the probability $p_{i'j',S}$ to zero for all i' , j' , and S where $j \notin S$. Similarly, if a link ij is forbidden from appearing, we can force our solution to obey this restriction by setting the probability $p_{ij,S}$ of the link to zero for all S .

Costly links. In our model, the optimal hyperlink structure of a web site depends on the transition probabilities of a links which in turn depend on the set of other links on the page. In addition, one could imagine a model in which each link ij incurs an associated cost c_{ij} . This situation can easily be handled in all our results by appropriately adjusting the maximization objective.

Location-dependent probabilities. We can model situations where page i has k_i “slots” for placing links to other pages, and the probability that a link is clicked also depends on the slot in which the link is placed.

Accounting for revisits. Our model thus far ignored the history of a surfer in defining the transition probabilities. However, in some settings it is reasonable

to assume that a surfer is less likely to return to a page he has already visited, especially in the recent past. The limiting case, when a surfer never returns to a page he has already visited, is NP-hard as can be seen easily by a reduction from the longest-path problem. Approximating this instance remains open, as does the computability of very interesting special cases of limited memory or simple history-dependent probability structures. In another model, we could assume that the transition probabilities remain the same regardless of the history, but the revenue structure changes, i.e., a surfer does not incur any extra revenue the second time he visits a page.

Handling different demographics. It is commonly acknowledged that different demographics have different surfing and purchasing patterns. A 21 year-old computer scientist from Seattle is more likely to navigate to the automotive section of Amazon.com than a 10 year-old school-girl from Wichita and is more likely to more spend money there than a 46 year-old farmer from Boise. One way to optimize a web site given such information is to dynamically update the link structure for each demographic, and indeed some of the larger web sites are starting to take this approach with a subset of their links. If dynamic links are an option for a web site, our results apply trivially by solving the problem separately for each demographic. However, static link structures are still the most prevalent style, and computing an optimal static link structure given demographic data and associated probabilities and revenues remains an interesting open problem.

The NTU game. As the proof of Scarf’s theorem uses an exponential-time algorithm (or, alternatively, fixed-point theorems), our result regarding existence of the core in the NTU game is non-constructive and we do not know how to find a fractional graph in the core in polynomial time. Furthermore, we do not know how to find an *integral* graph in the core or even prove that one always exists (although we have not been able to find a counter-example). It might be possible to prove existence of (and perhaps even compute) an integral graph in the core using potential proofs similar to those for proving existence of pure Nash equilibria in non-cooperative games.

The PageRank objective. One of the most commonly used systems for sorting web pages in search engine results is PageRank. As search engines are the single most essential portal to the web for most surfers, “search engine optimization” (SEO, as the industry calls it) of a web site is crucial to its success, so crucial that the commodity of PageRank sells for nearly \$100 on eBay⁵. Informally, the PageRank of a web page is defined as the probability of that page in the stationary distribution of a random walk on the web. Although the internal hyperlink structure of a web site does not affect it’s average (over all pages) PageRank, it does affect the maximum: to maximize the PageRank over all pages of a page in the web site, all pages should link to the page with highest entrance probability. This structure is trivial and unlikely to work given the search engine industry’s spam detection efforts. However, one could try to

⁵ Merchants selling PageRank are purportedly themselves high-ranked pages and are selling the placement of a text-link on their page with the sole intent of boosting the linked-to page’s rank.

maximize this objective with certain restrictions on the hyperlink structure that attempt to avoid detection like maximum in-degree. The NTU game also poses an interesting question.

Acknowledgments. We would like to thank Gary Flake for asking the question that motivated this work. We have benefitted from discussions with Christian Borgs, Jennifer Chayes, Uri Feige, and Gary Flake.

References

1. N. Badasyan and S. Chakrabarti. Private peering among internet backbone providers. mimeo: Virginia Tech., 2003.
2. O. N. Bondareva. Some applications of linear programming to cooperative games. *Problemy Kibernetiki*, 10:119–139, 1963.
3. T. Furusawa and H. Konishi. Free trade networks. mimeo: Yokohama National University and Boston College, 2002.
4. M. Granovetter. *Getting a job: a study of contacts and careers*. University of Chicago Press, 2 edition, 1995.
5. M. O. Jackson. A survey of models of network formation: Stability and efficiency. In G. Demange and M. Wooders, editors, *Group Formation in Economics; Networks, Clubs and Coalitions*, chapter 1, pages 11–57. Cambridge University Press, 2004.
6. M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.
7. M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, 1994.
8. H. E. Scarf. The core of an n-person game. *Econometrica*, 35:50–69, 1967.
9. L. S. Shapley. On balanced sets and cores. *Naval Research Logistics Quarterly*, 14:453–460, 1967.
10. J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*, chapter 13, pages 257–275. John Wiley and Sons, 1944.