JOURNAL OF ECONOMIC THEORY 76, 322–344 (1997)
ARTICLE NO. ET972306

Stable Networks*

Bhaskar Dutta and Suresh Mutuswami

Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi 110 016, India

Received November 13, 1995; revised February 24, 1997

A network is a graph where the nodes represent agents and an arc exists between two nodes if the corresponding agents interact bilaterally. An exogeneous value function gives the value of each network, while an allocation rule describes how the value is distributed amongst the agents. M. Jackson and A. Wolinsky (1996, *J. Econ. Theory* **71**, 44–74) have recently demonstrated a potential conflict between stability and efficiency in this framework. In this paper, we use an implementation approach to see whether the tension between stability and efficiency can be resolved. *Journal of Economic Literature* Classification Numbers: C72, D20.

1. INTRODUCTION

The interaction between agents can often be fruitfully described by a *network structure* or *graph*, where the nodes represent the agents and an arc exists between two nodes if the corresponding agents interact bilaterally. Network structures have been used in a wide variety of contexts ranging from social networks (Wellman and Berkowitz [16]), information transmission (Goyal [4]), internal organization of firms (Marschak and Reichelstein [10]), cost allocation schemes (Henriet and Moulin [7]), to the structure of airline routes (Hendricks *et al.* [6]).[1]

In a recent paper, Jackson and Wolinsky [8] focus on the *stability* of networks. Their analysis is designed to give predictions concerning which networks are likely to form when self-interested agents can choose to form new links or severe existing links. They use a specification where a *value function* gives the value (or total product) of each graph or network, while an allocation rule gives the distribution of value amongst the agents forming the network. A principal result of their analysis shows that efficient

* We are most grateful to Matt Jackson for several helpful discussions and suggestions. Thanks are also due to Sudipto Bhattacharyya, an anonymous referee, and an Associate Editor for comments on earlier versions of the paper. An earlier version of the paper was written when the first author was visiting Caltech. Their hospitality is gratefully acknowledged.

[1] See van den Nouweland [13] and Sharkey [15] for detailed surveys and additional references.

graphs (that is, graphs of maximum value) may not be stable when the allocation rule treats individuals symmetrically.

The main purpose of this paper is to subject the potential conflict between stability and efficiency of graphs to further scrutiny. In order to do this, we follow Dutta *et al.* [3] and assume that agents' decisions on whether or not to form a link with other agents can be represented as a game in strategic form.[2] In this "link formation" game, each player announces a set of players with whom he or she wants to form a link. A link between two players is formed if both players want the link. This rule determines the graph corresponding to any n-tuple of announcements. The value function and the allocation rule then give the payoff function of the strategic form game.

Since the link formation game is a well-defined strategic form game, one can use any equilibrium concept to analyze the formation of networks. In this paper, we will define a graph to be *strongly stable* (respectively *weakly stable*) if it corresponds to a *strong Nash* equilibrium (respectively *coalition-proof Nash equilibrium*) of the link formation game. Although Jackson and Wolinsky [8] did not use the link formation game, their specification assumed that only *two*-person coalitions can form; their notion of *pairwise stability* is implied by our concept of strong stability. Hence, it follows straightaway from their analysis that there is a conflict between strong stability and efficiency if the allocation rule is symmetric.

How can we ensure that efficient graphs will form? One possibility is to use allocation rules which are *not* symmetric. For instance, fix a vector of weights $w = (w_1, w_2, ..., w_n)$. Call an allocation rule *w-fair* if the gains or losses to players i and j from the formation of the *new* link (ij) is proportional to w_i/w_j. *w-fair* rules are symmetric only if $w_i = w_j$ for all i and j. However, the vector of weights w can be chosen so that there is only a "slight" departure from symmetry. We first show that the class of w-fair rules coincides with the class of *weighted Shapley values* of an appropriately defined transferable utility game. We then go on to construct a value function under which *no* efficient graph is strongly stable for any w-fair allocation rule. Thus, the relaxation of symmetry in this direction does not help.

A second possibility is to use weak stability instead of strong stability. However, again we demonstrate a conflict between efficiency, symmetry and (weak) stability.

We then go on to adopt an *implementation* or *mechanism design* approach. Suppose the implicit assumption or prediction is that only those graphs

[2] This game was originally suggested by Myerson [12] and subsequently used by Qin [14]. See also Hart and Kurz [5] who discuss a similar strategic form game in the context of thy endogeneous formation of coalition structures.

which correspond to strong Nash equilibria of the link formation game will form. Then, our interest in the *ethical* properties of the allocation rule should be restricted only to how the rule behaves *on the class of these graphs*. Hence, if we want symmetry of the allocation rule, we should be satisfied if the allocation rule is symmetric on the *subdomain* of strongly stable graphs.

We analyse *two* specific problems within this general approach. In the first *design* problem, we construct an allocation rule which ensures that (i) the class of strongly stable graphs is a *nonempty* subset of the set of efficient graphs, and (ii) satisfies the restriction that the rule is symmetric on the class of strongly stable graphs. This result is proved under a very mild restriction on the class of value functions. The second result is much stronger, but is proved for a *more* restrictive class of value functions. More specifically, we construct an allocation rule which (given the restrictions on the class of value functions) guarantees that (i) there is at least *one* strongly stable graph, (ii) all *weakly stable* graphs are efficient, and (iii) the allocation rule is symmetric on the class of weakly stable graphs. Thus, this achieves a kind of "double" implementation in strong Nash and coalition-proof Nash equilibrium.

A common feature of the allocation rules constructed by us is that these distribute the value of stable graphs *equally* amongst all agents. Obviously, this ensures symmetry of the allocation rules on the class of stable graphs. Of course, the rules do not treat agents symmetrically on some graphs which are not stable. Indeed, the asymmetries are carefully constructed so as to ensure that the other requirements of the design problem(s) are satisfied.

The plan of this paper is as follows. In Section 2, we provide definitions of some key concepts. Section 3 describes the link formation game, while Section 4 contains the results. We conclude in Section 5.

2. SOME DEFINITIONS

Let $N=\{1, 2, ..., n\}$ be a finite set of agents with $n \geqslant 3$. Interactions between agents are represented by graphs whose vertices represent the players, and whose arcs depict the pairwise relations. The *complete* graph, denoted g^N, is the set of *all* subsets of N of size 2. G is the set of *all* possible graphs on N, so that $G=\{g \mid g \subset g^N\}$.

Given any $g \in G$, let $N(g)=\{i \in N \mid \exists j$ such that $(ij) \in g\}$.

The link (ij) is the subset of N containing i, j. $g+(ij)$ and $g-(ij)$ are the graphs obtained from g by adding and subtracting the link (ij) respectively.

i and j are *connected* in g if there is a sequence $\{i_0, i_1, ..., i_K\}$ such that $i_0=i$, $i_K=j$ and $(i_k i_{k+1}) \in g$ for all $k=0, 1, ..., K-1$. We will use $C(g)$ to denote the set of connected components of g. g is said to be *fully connected* (respectively *connected* on S) if all pairs of agents in N (respectively in S) are connected. g is *totally disconnected* if $g = \{\varnothing\}$. If h is a component of g, then $N(h) = \{i \mid (ij) \in h \text{ for some } j \in N \backslash \{i\}\}$, and n_h denotes the cardinality of $N(h)$.

The value of a graph is represented by a function $v: G \to \mathbb{R}$. We will only be interested in the set V of such functions satisfying *Component Additivity*.

DEFINITION 2.1. A value function is component additive if $v(g) = \sum_{h \in C(g)} v(h)$.

We interpret the value function to indicate the total "output" produced by agents in N when they are "organized" according to a particular graph. For instance, the members of N may be workers in a firm. The graph g then represents the internal organization of the firm, that is the structure of communication amongst the workers. Alternatively, N could be a set of (tax) auditors and supervisors, and g could represent a particular hierarchical structure of auditors and supervisors. In this case, $v(g)$ is the (expected) tax revenue realized from a population of tax payers when g is in "operation."

It is worth emphasizing at this point that the value function is a very general concept. In particular, it is more general than Myerson's [11] games with cooperation structure. A cooperation structure is a graph in our terminology. Given any exogeneously specified transferable utility game (N, u) and a graph g, we define for each $S \subset N$, the restricted graph on S as $g \mid S \equiv \{(ij) \in g \mid i, j \in S\}$. The *graph-restricted* game (N, u^g) specifies the *worth* of a coalition as follows.

$$\text{For all} \quad S \subset N, u^g(S) = \sum_{h \in C(g \mid S)} u(N(h)). \tag{2.1}$$

As (2.1) makes clear, the value or worth of a given set of agents in Myerson's formulation depends on whether they are connected or not, whereas in the Jackson–Wolinsky approach, the value of a coalition can in principle depend on *how* they are connected.

Given v, g is *strongly efficient* if $v(g) \geqslant v(g')$ for all $g' \in G$. Let $E(v)$ denote the set of strongly efficient graphs.

Finally, an *allocation rule* $Y: V \times G \to \mathbb{R}^N$ describes how the value associated with each network is distributed to the individual players. $Y_i(v, g)$ will denote the payoff to player i from graph g under the value function v. Clearly, an allocation rule corresponds to the concept of a *solution* in cooperative game theory.

Given a permutation $\pi: N \to N$, let $g^\pi = \{(ij) \mid i = \pi(k), j = \pi(l), (kl) \in g\}$. Let v^π be defined by $v^\pi(g^\pi) = v(g)$.

The following condition imposes the restriction that all agents should be treated symmetrically by the allocation rule. In particular, *names* of the agents should not determine their allocation.

DEFINITION 2.2. Y is *anonymous* on $G' \subseteq G$ if for *all* pairs $(v, g) \in V \times G'$, and for all permutations π, $Y_{\pi(i)}(v^{\pi}, g^{\pi}) = Y_i(v, g)$.

Remark 2.3. If Y is anonymous on G, we say that Y is *fully anonymous.*

DEFINITION 2.4. Y is *component balanced* if $\sum_{i \in N(h)} Y_i(v, g) = v(h)$ for every $g \in G$, $h \in C(g)$.

Component balance implies that cross-subsidization is ruled out. We will restrict attention to component balanced allocation rules throughout the paper.[3]

3. THE LINK FORMATION GAME

In this section, we describe the strategic form game which will be used to model the endogenous formation of networks or graphs.[4] The following description of the link formation game assumes a specific value function v and an allocation rule Y. Let $\gamma \equiv (v, Y)$.

The linking game $\Gamma(\gamma)$ is given by the $(n+2)$-tuple $(N; S_1, ..., S_n, f^{\gamma})$, where for each $i \in N$, S_i is player i's *strategy* set with $S_i = 2^{N \setminus \{i\}}$ and the *payoff function* is the mapping $f^{\gamma}: S \equiv \prod_{i \in N} S_i \Rightarrow \mathbb{R}^N$ given by

$$f_i^{\gamma}(s) = Y_i(v, g(s)) \tag{3.1}$$

for all $s \in S$, with

$$g(s) = \{(ij) \mid j \in s_i, i \in s_j\}. \tag{3.2}$$

So, a typical strategy of player i in $\Gamma(\gamma)$ consists of the set of players with whom i wants to form a link. Then, (3.2) states that a link between i and j forms if and only if they both want to form this link. Hence, each strategy vector gives rise to a unique graph $g(s)$. Finally, the payoff to player i associated with s is simply $Y_i(v, g(s))$, the payoff that, is given by the allocation rule for the graph *induced by s*.[5]

[3] Jackson and Wolinsky [8] point out that the conflict between anonymity, stability, and efficiency disappears if the rule is not component balanced.

[4] Aumann and Myerson [1] use an *extensive* form approach in modeling the endogeneous formation of cooperation structures.

[5] We will say that g is induced by s if $g \equiv g(s)$, where $g(s)$ satisfies (3.2).

We now define some equilibrium concepts for $\Gamma(\gamma)$.

DEFINITION 3.1. A strategy vector $s^* \in S$ is a *strong Nash equilibrium* (SNE) of $\Gamma(\gamma)$ if there is *no* $T \subseteq N$ and $s \in S$ such that

(i) $s_i = s_i^*$ for all $i \notin T$.

(ii) $f_i^\gamma(s) > f_i^\gamma(s^*)$ for all $i \in T$.

The second equilibrium concept that will be used in this paper is that of coalition-proof Nash equilibrium (CPNE). In order to define the concept of CPNE of $\Gamma(\gamma)$, we need some more notation. For any $T \subset N$, and $s^*_{N|T} \in S_{N|T} \equiv \prod_{i \in N|T} S_i$, let $\Gamma(\gamma, s^*_{N|T})$ denote the game induced on T by $s^*_{N|T}$. So,

$$\Gamma(\gamma, s^*_{N|T}) = \langle T, \{S_i\}_{i \in T}, \bar{f}^\gamma \rangle, \tag{3.3}$$

where for all $j \in T$, for all $s_T \in S_T$, $\bar{f}_j^\gamma(s_T) = f_j^\gamma(s_T, s^*_{N|T})$.

The set of CPNE of $\Gamma(\gamma)$ is defined inductively on the set of players.

DEFINITION 3.2. In a single-player game, s^* is a CPNE of $\Gamma(\gamma)$ iff s_i^* maximises $f_i^\gamma(s)$ over S. Let $\Gamma(\gamma)$ be a game with n players, where $n > 1$. Suppose CPNE have been defined for all games with less than n players. Then, (i) $s^* \in S$ is *self-enforcing* if for all $T \subset N$, s_T^* is a CPNE of $\Gamma(\gamma, s^*_{N|T})$; and (ii) $s^* \in S$ is a CPNE of $\Gamma(\gamma)$ if it is self-enforcing and moreover there does not exist another self-enforcing strategy vector $s \in S$ such that $f_i^\gamma(s) > f_i^\gamma(s^*)$ for all $i \in N$.

Our interest lies not in the strategy vectors which are SNE or CPNE of $\Gamma(\gamma)$, but in the *graphs* which are induced by these equilibria. This motivates the following definition.

DEFINITION 3.3. g^* is *strongly stable* [respectively *weakly stable*] for $\gamma = (v, Y)$ if g^* is induced by some s which is a SNE [respectively CPNE] of $\Gamma(\gamma)$.

Hence, a strongly stable graph is induced or supported by a strategy vector which is a strong Nash equilibrium of the linking game. Of course, a strongly stable graph must also be weakly stable.

Finally, in order to compare the Jackson–Wolinsky notion of *pairwise stability*, suppose the following constraints are imposed on the set of possible deviations in Definition 3.1. First, the deviating coalition can contain at most *two* agents. Second, the deviation can consist of severing just *one* existing link or forming *one additional* link. Then, the set of graphs which are immune to such deviations is called pairwise stable. Obviously, if g^* is strongly stable, then it must be pairwise stable.

4. THE RESULTS

Notice that strong stability (as well as weak stability) has been defined for a *specific* value function v and allocation rule Y. Of course, which network structure is likely to form must depend upon both the value function as well as on the allocation rule. Here, we adopt the approach that the value function is given *exogeneously*, while the allocation rule itself can be "chosen" or "designed."

Within this general approach, it is natural to seek to construct allocation rules which are (ethically) attractive and which also lead to the formation of stable network structures which maximize output, no matter what the exogeneously specified value function. This is presumably the underlying motive behind Jackson and Wolinsky's search for a symmetric allocation rule under which at least *one* strongly efficient graph would be pairwise stable for every value function.

Given their negative result, we initially impose weaker requirements. First, instead of *full anonymity*, we only ask that the allocation rule be *w-fair*, a condition which is defined presently. However, we show that there can be value functions under which *no* strongly efficient graph is strongly stable.[6] Second, we retain full anonymity but replace strong stability by weak stability. Again, we construct a value function under which the unique strongly efficient graph is *not* weakly stable.

Our final results, which are the main results of the paper, explicitly adopt an implementation approach to the problem. Assuming that strong Nash equilibrium is the "appropriate" concept of equilibrium and that the individual agents decide to form network relations through the link formation game is equivalent to predicting that only strongly stable graphs will form. Let $S(\gamma)$ be the set of strongly stable graphs corresponding to $\gamma \equiv (v, Y)$. Instead of imposing full anonymity, we only require that the allocation rule be anonymous on the *restricted* domain $S(\gamma)$. However, we now require that for all permissible value functions, $S(\gamma)$ is contained in the set of strongly efficient graphs, instead of merely *intersecting* with it, which was the "target" sought to be achieved in the earlier results. We are able to construct an allocation rule which satisfies these requirements.

Suppose, however, that the designer has some doubt whether strong Nash equilibrium is really the "appropriate" notion of equilibrium. In particular, she apprehends that weakly stable graphs may also form. Then, she would want to ensure anonymity of the allocation rule over the larger class of weakly stable graphs, as well as efficiency of these graphs. Assuming a stronger restriction on the class of permissible value functions, we are able to construct an allocation rule which satisfies these requirements. In

[6] We point out below that strong stability can be replaced by pairwise stability.

addition the allocation rule also guarantees that the set of strongly stable graphs is nonempty.

Our first result uses w-fairness. Fix a vector $w=(w_1, ..., w_n) \gg 0$.

DEFINITION 4.1. An allocation rule Y is *w-fair* if for all $v \in V$, for all $g \in G$, for all $i, j \in N$,

$$\frac{1}{w_i}[Y_i(v, g) - Y_i(v, g-(ij))] = \frac{1}{w_j}[Y_j(v, g) - Y_j(v, g-(ij))].$$

In Proposition 4.1 below, we show that *the* unique allocation rule which satisfies w-fairness and component balance is the *weighted Shapley value* of the following characteristic function game.

Take any $(v, g) \in V \times G$. Recall that for any $S \subseteq N$, the restricted graph on S is denoted $g|S$. Then, the TU game $u_{v,g}$ is given by:

For all $S \subseteq N$, $u_{v,g}(S) = \sum_{h \in C(g|S)} v(h)$.

PROPOSITION 4.2. *For all $v \in V$, the unique w-fair allocation rule Y which satisfies components balance is the weighted Shapley value of $u_{v,g}$.*

Proof. The proof is omitted since it is a straightforward extension of the corresponding result in Dutta *et al.* [3].

Remark 4.3. This proposition is similar to corresponding results of Dutta *et al.* [3] and Jackson and Wolinsky [8]. The former proved that w-fair allocation rules satisfying component balance are the weighted Shapley values (also called weighted Myerson values) of the *graph-restricted* game given by any exogeneous TU game and any graph g. Of course, the set of graph-restricted games is a strict subset of V, and hence Proposition 4.2 is a formal generalization of their result. Jackson and Wolinsky show that where $w_i = w_j$ for all i, j, then the unique w-fair allocation rule satisfying component balance is the Shapley value of $u_{v,g}$.

Our first result on stability follows. The motivation for proving this result is the following. Since the weight vector w can be chosen to make the allocation rule "approximately" anonymous (by choosing w to be very close to the unit vector $(1, ..., 1)$, we may "almost" resolve the tension between stability, efficiency and symmetry unearthed by Jackson–Wolinsky by using such a w-fair allocation rule. However, the next result rules out this possibility.

THEOREM 4.4. *Suppose $w \gg 0$. Then, there is no w-fair allocation rule Y satisfying component balance and such that for each $v \in V$, at least one strongly efficient graph is strongly stable.*

Proof. Let $N=\{1, 2, 3\}$, and choose any w such that $w_1 \geqslant w_2 \geqslant w_3 > 0$ and $\sum_{i=1}^{3} w_i = 1$.

Now, consider the (component additive) v such that $v(\{(ij)\})=1$, $v(\{(ij), (jk)\})=1+\varepsilon$, and $v(g^N)=1+2\varepsilon$, where $\varepsilon \in (0, \frac{1}{2}(1-(w_2/(w_1+w_3))))$.

Using Proposition 4.2, the unique w-fair allocation rule Y satisfying component balance is the weighted Shapley value of $u_{v,g}$. Routine calculation yields

$$Y_i(v, g^N) = 2\varepsilon w_i + w_i \left(\frac{w_k}{w_i + w_j} + \frac{w_j}{w_i + w_k} \right) \qquad \text{for all} \quad i, j, k \in N \tag{4.1}$$

$$Y_i(v, \{(ij)\}) = \frac{w_i}{w_i + w_j} \qquad \text{for all} \quad i, j \in N. \tag{4.2}$$

From (4.1) and (4.2), and using $\sum_{i=1}^{3} w_i = 1$

$$Y_i(v, \{(ij)\}) - Y_i(v, g^N) = w_i \left(1 - 2\varepsilon - \frac{w_j}{w_i + w_k} \right). \tag{4.3}$$

Remembering that $w_1 \geqslant w_2 \geqslant w_3$ and that $\varepsilon < \frac{1}{2}(1-(w_2/(w_1+w_3))))$, (4.3) yields

$$Y_i(v, \{(ij)\}) - Y_i(v, g^N) > 0 \qquad \text{for} \quad i \in \{2, 3\} \tag{4.4}$$

which implies that g^N is not strongly stable since $\{2, 3\}$ will break links with 1 to move to the graph $\{(2, 3)\}$. Since g^N is the unique strongly efficient graph, the theorem follows.

Remark 4.5. Note that since only a pair of agents need form a coalition to "block" g^N, the result strengthens the intuitive content of the Jackson–Wolinsky result.

Our next result uses weak stability instead of strong stability.

THEOREM 4.6. *There is no fully anonymous allocation rule Y satisfying component balance such that for each $v \in V$, at least one strongly efficient graph is weakly stable.*

Proof. Let $N=\{1, 2, 3\}$, and consider v such that $v(g^N)=1=v(\{(ij)\})$ and $v(\{(ij), (jk)\})=1+2\varepsilon$. Assume that $0<\varepsilon<\frac{1}{12}$.

Since Y is fully anonymous and component balanced, $Y_i(v, \{(ij)\}) = Y_j(v, \{(ij)\}) = \frac{1}{2}$. Let $g^j \equiv \{(ij), (jk)\}$. Note that $\{g^j \mid j \in N\}$ is the set of strongly efficient graphs. Choose any $j \in N$. Then, $Y_j(v, g^j) \geqslant \frac{1}{2}$. For, suppose $Y_j(v, g^j) < \frac{1}{2}$. Then, j can deviate *unilaterally* to change g^j to $\{(ij)\}$ or $\{(jk)\}$ by breaking the link with i or k respectively. So, if $Y_j(v, g^j) < \frac{1}{2}$

and g^j is induced by s, then s is not a Nash equilibrium, and hence not a CPNE.

So, $Y_j(v, g^j) \geqslant \frac{1}{2}$. Since Y is fully anonymous and component balanced, $Y_i(v, g^j) = Y_k(v, g^j) \leqslant \frac{1}{4} + \varepsilon$. Again, full anonymity of Y ensures that $Y_i(v, g^N) = \frac{1}{3}$ for all $i \in N$.

Hence, $\{i, k\}$ can deviate from g^j and form the additional link (ik). This will precipitate the complete graph. From preceding arguments, the deviation is profitable if $\varepsilon < \frac{1}{12}$.

Letting s^N denote the strategy n-tuple which induces g^N, one notes that s^N is a Nash equilibrium. Hence, the deviation of $\{i, k\}$ to s^N is not deterred by the possibility of a further deviation by either i or k. So, g^j is *not* weakly stable. This completes the proof of the theorem.

Remark 4.7. Again note that only a 2-person coalition has to form to block g^j. So, the result could have been proved in terms of "pairwise weak stability," which is strictly weaker than pairwise stability. Hence, this generalizes Jackson and Wolinsky's basic result.

DEFINITION 4.8. v satisfies *monotonicity* if for all $g \in G$, for all $i, j \in N$, $v(g + (ij)) \geqslant v(g)$.

Thus, a monotonic value function has the property that additional links never decrease the value of the graph.[7] A special class of monotonic value functions, the class considered by Dutta *et al.* [3], is the set of graph-restricted games derived from superadditive TU games. Of course, there are also other contexts which might give rise to monotonic value functions. Dutta *et al.* proved that for the class of graph-restricted games derived from superadditive TU games, a large class of component balanced allocation rules (including all w-fair rules with $w \gg 0$) has the property that the set of weakly stable graphs is a *subset* of the set of graphs which are *payoff-equivalent* to g^N. Moreover, g^N itself is weakly stable.[8] Their proof can be easily extended to cover *all* monotonic value functions. We state the following result.

THEOREM 4.9. *Suppose v is monotonic. Let Y be any w-fair allocation rule with $w \gg 0$, and satisfying component balance. Then, g^N is weakly stable for (v, Y). Moreover, if g is weakly stable for (v, Y), then g is payoff-equivalent to g^N.*[9]

[7] Hence, g^N is strongly efficient.

[8] g and g' are payoff-equivalent under (v, Y) if $Y(v, g) = Y(v, g')$. Also, note that if v is monotonic, then g^N and hence graphs which are payoff-equivalent, to g^N are strongly efficient.

[9] The result is true for a larger class of allocation rules, which is not being defined here to save space.

The result is true for a larger class of allocation rules, which is not being defined here to save space.

Proof. The proof is omitted since it is almost identical to that of Dutta *et al.* [3].

Remark 4.10. Note that Theorem 4.9 ensures that *only* strongly efficient graphs are weakly stable. Thus, if our prediction is that only weakly stable graphs will form, then this result guarantees that there will be no loss in efficiency. This guarantee is obviously stronger than that provided if *some* stable graphs is strongly efficient. In the latter case, there is the possibility that *other* stable graphs are inefficient, and since there is no reason to predict that *only* the efficient stable graph will form, inefficiency can still occur.

Unfortunately, monotonicity of the value function is a stringent requirement. There are a variety of problems in which the optimum network is a *tree* or a *ring*. For example, cost allocation problems give rise to the minimum-cost spanning tree. Efficient airline routing or optimal trading arrangements may also imply that the *star* or *ring* is the efficient network.[10] Indeed, in cases where there is a (physical) cost involved in setting up an additional link, g^N will seldom be the optimal network.

This provides the motivation to follow the "implementation approach" and prove results similar to that of Theorem 4.9, but covering *nonmonotonic* value functions. First, we construct a component balanced allocation rule which is anonymous on the set of *strongly stable graphs* and which ensures that *all* strongly stable graphs are strongly efficient.

In order to prove this result, we impose a restriction on the class of value functions.

DEFINITION 4.11. The set of *admissible* value functions is $V^* = \{v \in V \mid v(g) > 0 \text{ iff } g \text{ is not totally disconnected}\}$.

So, a value function is admissible if *all* graphs (except the trivial one in which no pair of agents is connected) have positive value.[11]

Before we formally define the particular allocation rule which will be used in the proof of the next theorem, we discuss briefly the key properties which will have to be satisfied by the rule.

[10] See Hendricks *et al.* [6] on the "hub" and "spokes" model of airline routing. See also Landa [9] for an interesting account of why the ring is the efficient institutional arrangement for organization of exchange amongst tribal communities in East Papua New Guinea.

[11] In common with Jackson and Wolinsky, we are implicitly assuming that the value of a disconnected player is zero. This assumption can be dropped at the cost of some complicated notation.

Choose some efficient $g^* \in G$. Suppose s^* induces g^*, and we want to ensure that g^* is strongly stable. Now, consider any g which is different from g^*, and let s induce g. Then, the allocation rule must *punish* at least one agent who has deviated from s^* to s. This is possible only if a deviant can be *identified*. This is trivial if there is some $(ij) \in g \backslash g^*$, because then both i and j must concur in forming the extra link (ij). However, if $g \subset g^*$, say $(ij) \in g^* \backslash g$, then *either i or j* can unilaterally break the link. The only way to ensure that the deviant is punished, is to punish *both* i and j.

Several simple punishment schemes can be devised to ensure that at least two agents who have deviated from s^* are punished sufficiently to make the deviation unprofitable. However, since the allocation rule has to be component balanced, these punishment schemes may result in some other agent being given more than the agent gets in g^*. This possibility creates a complication because the punishment scheme has to *satisfy an additional property*. Since we also want to ensure that inefficient graphs are *not* strongly stable, agents have to be provided with an incentive to deviate from any graph which is not strongly efficient. Hence, the punishment scheme has to be relatively more sophisticated.

Choose some strongly efficient g^* with $C(g^*) = \{h_1^*, ..., h_J^*\}$, and let $\succ$ be a *strict* ordering on arcs of g^*. Consider any other graph g, and let $C(g) = \{h_1, ..., h_K\}$.

The first step in the construction of the allocation rule is to choose agents who will be punished in some components $h_k \in C(g)$. For reasons which will become clear later on, we only need to worry about components h_k such that $D(h_k) = \{i \in N(h_k) \mid (ij) \in g^* \text{ for some } j \notin N(h_k)\}$ is nonempty. For such components, choose $i(h_k) \equiv i_k$ such that $\forall j \in N(h_k) \backslash \{i_k\}$, $\forall m \notin N(h_k)$:

$$(jm) \in g^* \Rightarrow (i_k l) \succ (jm) \qquad \text{for some} \quad (i_k l) \in g^*, l \notin N(h_k). \qquad (4.5)$$

We will say that g is a $*$-*supergraph* of g^* if for each $h^* \in C(g^*)$, there is $h \in C(g)$ such that $N(h^*) \subseteq N(h)$. Note that the fully connected graph is a $*$-*supergraph* of every graph.

LEMMA 4.12. *Suppose $g = (h_1, ..., h_K)$ is not a $*$-supergraph of g^*. Then, $\exists k, l \in \{1, ..., K\}$ such that $(i_k i_l) \in g^*$.*

Proof. Since g is not a $*$-*supergraph*, it follows that g is not fully connected, and that there exists a component h and players i, j such that $i \in N(h)$, $j \notin N(h)$ and $(ij) \in g^*$. Indeed, assume that for each $h_k \in C(g)$, the set $D(h_k)$ is nonempty.[12] For every $k = 1, 2, ..., K$, there is $j_k \notin N(h_k)$ such that $(i_k j_k) \in g^*$ and $(i_k j_k) \succ (ij)$ for all $i \in D(h_k) \backslash \{i_k\}$ and for all $j \notin N(h_k)$ with $(ij) \in g^*$. Let the $\succ$-maximal element within the set $\{(i_1, j_1), ..., (i_K j_K)\}$ be

[12] Otherwise, we can restrict attention to those components for which $D(h_k)$ is nonempty.

$(i_{k^*}\, j_{k^*})$. Let $j_{k^*} \in N(h_l)$. Note that from the definition of the pair $(i_{k^*}\, j_{k^*})$, it follows that $l \neq k^*$. Also, $(i_{k^*}\, j_{k^*}) \in g^*$. It therefore follows that $i_l = j_{k^*}$ and $j_l = i_{k^*}$. Hence, $(i_{k^*} i_l) \in g^*$. This completes the proof of the lemma.

The implication of Lemma 4.12 is the following. Suppose one or more agents deviate from g^* to some $g \in G$ with components $\{h_1, ..., h_K\}$. Then, the set of agents $\{i(h_1), ..., i(h_K)\}$ must contain a deviator. This property will be used intensively in the proof of the next theorem.

THEOREM 4.13. *Let $v \in V^*$. Then, there is a component balanced allocation rule Y^* such that the set of strongly stable graphs is nonempty and contained in $E(v)$. Moreover, Y^* is anonymous on the set of strongly stable graphs.*

Proof. Choose any $v \in V^*$. Fix $g^* \in E(v)$. Let $C(g^*) = \{h_1^*, ..., h_K^*\}$. An allocation rule Y^* satisfying the required properties will be constructed which ensures that g^* is strongly stable. Moreover, no g will be strongly stable unless it is in $E(v)$.

For any $S \subseteq N$ with $|S| \geqslant 2$, let G_S be the set of graphs which are connected on S, and have no arcs outside S. So,

$$G_S = \{g \in G \mid g \text{ is connected on } S \text{ and } N(g) = S\}.$$

Let

$$a_S = \min_{g \in G_S} \frac{v(g)}{|S|(n-2)}.$$

Choose any ε such that

$$0 < \varepsilon < \min_{S \subseteq N} a_S \tag{4.6}$$

The allocation rule Y^* is defined by the following rules. Choose any g.

(Rule 1) For any $h \in C(g)$, suppose $N(h) = \bigcup_{i \in I} N(h_i^*)$ for some (non-empty) $I \subseteq \{1, ..., K\}$. Then,

$$Y_i^*(v, g) = \frac{v(h)}{n_h} \qquad \text{for all} \quad i \in N(h).$$

(Rule 2) Suppose $N(h) \neq \bigcup_{i \in I} N(h_i^*)\ \forall I \subseteq \{1, ..., K\}$. Then, g is not a *$*$-supergraph* of g^*. Choose $j_h \in N(h)$ such that $j_h \neq i_h$. Then,

$$Y_i^*(v, g) = \begin{cases} (n_h - 1)\varepsilon & \text{if} \quad i \neq j_h \\ v(h) - (n_h - 1)^2\, \varepsilon & \text{otherwise.} \end{cases}$$

Clearly, the rule defined above is component balanced. We will show later that Y^* is anonymous on the set of strongly stable graphs. We first show that the efficient graph g^* is strongly stable under the above allocation rule.

Let s^* be the strategy profile defined as follows: For all $i \in N$, $s_i^* = \{j \in N \mid (ij) \in g^*\}$. Clearly, s^* induces g^* in $\gamma = (v, Y^*)$. We need to show that s^* is a SNE of $\Gamma(\gamma)$.

Consider any $s \neq s^*$, and let g be induced by s. Also, let $T = \{i \in N \mid s_i \neq s_i^*\}$. Suppose $h \in C(g)$. If $N(h) = \bigcup_{i \in I} N(h_i^*)$ for some nonempty subset I of $\{1, ..., K\}$, then $Y_i^*(v, g) = v(h)/n_h$ for all $i \in N(h)$. However, since g^* is efficient, there exists some $i \in I$ such that $v(h_i^*)/n_{h_i^*} \geqslant v(h)/n_h$. So, no member of h_i^* is better-off as a result of the deviation. Also, note that $T \cap N(h_i^*) \neq \emptyset$. So, T does not have a profitable deviation in this case.

Suppose there is $h \in C(g)$ such that $N(h) \neq \bigcup_{i \in I} N(h_i^*)$ for any nonempty subset I of $\{1, ..., K\}$. Then, g is not a $*$-*supergraph* of g^*, and let $C(g) = \{h_1, ..., h_L\}$. From the above lemma, there exists $(i_k i_l) \in g^*$ where i_k and i_l are the players who are punished in h_k and h_l respectively. Obviously, $T \cap \{i_k, i_l\} \neq \phi$. But from Rule (2), it follows that $Y_{i_k}^*(v, g) = (n_{h_k} - 1)\varepsilon$ and $Y_{i_l}^*(v, g) = (n_{h_l} - 1)\varepsilon$. Given the value of ε, it follows that both i_k and i_h are worse-off from the deviation.

We now show that if g is strongly stable, then $g \in E(v)$. So suppose that g is an inefficient graph.

(i) If g is an inefficient graph which is a $*$-*supergraph* of g^*, then there exist $h \in C(g)$, $h^* \in C(g^*)$ such that $N(h^*) \subseteq N(h)$ and

$$Y_i^*(v, g) = \frac{v(h)}{n_h} < Y_i^*(v, g^*) = \frac{v(h^*)}{n_{h^*}} \qquad \text{for all} \quad i \in N(h^*).$$

So, each $i \in N(h^*)$ can deviate to the strategy s_i^*. This will induce the component h^* where they are all strictly better off.

(ii) Suppose that g is not a $*$-*supergraph* of g^*. Let $C(g) = (h_1, ..., h_K)$. Without loss of generality, let $n_{h_1} \leqslant \cdots \leqslant n_{h_K}$. Since g is not a $*$-*supergraph* of g^*, Rule (2) of the allocation rule applies and we know that there exist $h_k, h_l \in C(g)$, and $i_{h_k} \in N(h_k)$, $i_{h_l} \in N(h_l)$ such that $Y_{i_{h_k}}^*(v, g) = (n_{h_k} - 1)\varepsilon$ and $Y_{i_{h_l}}^*(v, g) = (n_{h_l} - 1)\varepsilon$. Let $\bar{s}$ be such that

(i) $\forall j \notin \{i_{h_k}, i_{h_l}\}, \bar{s}_j = s_j$.

(ii) $\bar{s}_{i_{h_k}} = \{j \mid j \in s_{i_{h_k}} \text{ or } j = i_{h_l}\}$.

(iii) $\bar{s}_{i_{h_l}} = \{j \mid j \in s_{i_{h_l}} \text{ or } j = i_{h_k}\}$.

Let $\bar{g}$ be the graph induced by $\bar{s}$. Notice that $\bar{g} = g + (i_{h_k} i_{h_l})$. We claim that

$$Y_j^*(v, \bar{g}) > Y_j^*(v, g) \; \forall j \in \{i_{h_k}, i_{h_l}\}. \tag{4.7}$$

Let $\bar{h} \in C(\bar{g})$ be the component containing players i_{h_k} and i_{h_l}. Notice that $n_{\bar{h}} > \max(n_{h_k}, n_{h_l})$. Given the value of ε, it follows that

$$Y_j^*(v, \bar{g}) \geqslant (n_{\bar{h}} - 1)\,\varepsilon \qquad \forall j \in \{i_{h_k}, i_{h_l}\}.$$

This shows that the coalition $\{i_{h_k}, i_{h_l}\}$ has a deviation which makes both players better off.

The second half of the proof also shows that g is strongly stable only if g is a $*$-*supergraph* of g^*. From Rule (1), it is clear that Y^* is anonymous on all such graphs. This observation completes the proof of the theorem.

We have remarked earlier that we need to restrict the class of permissible value functions in order to prove the analogue of a "double implementation" in strong Nash and coalition-proof Nash equilibrium. In order to explain the motivation underlying the restricted class of value functions, we first show in Example 4.14 below that the allocation rule used in Theorem 4.13 *cannot* be used to prove the double implementation result. In particular, this allocation rule does not ensure that weakly stable graphs are efficient.

EXAMPLE 4.14. Let $N = \{1, 2, 3, 4\}$. Consider a value function such that $v(g^*) = 4$, $v(g_1) = 3.6$, $v(g_2) = v(g_3) = 2.9$, where $g^* = \{(14), (13), (23), (12)\}$, $g_1 = \{(12), (13), (34)\}$, $g_2 = \{(12), (13)\}$ and $g_3 = \{(13), (34)\}$. Also, $v(\{(ij)\}) = 1.6$. Finally, the value of other graphs is such that $\varepsilon = 0.4$ satisfies (4.6). Note that g^* is the unique efficient graph. Let the strict order on links (used in the construction of the allocation rule in Theorem 4.13) be given by

$$(13) \succ (23) \succ (14) \succ (12).$$

Consider the graph $g = \{(12), (34)\}$. Then, from (Rule 2) and the specification of $\succ$, we have $Y_2^*(v, g) = Y_4^*(v, g) = 1.2$, $Y_1^*(v, g) = Y_3^*(v, g) = 0.4$. Now, g is weakly stable, but not efficient.

To see that g is weakly stable, notice first that agents 2 and 4 have no profitable deviation. Second, check using the particular specification of $\succ$ that $Y_3^*(v, g_2) = 1.3 > Y_3^*(v, g_1) = 0.9$, $Y_1^*(v, \{(13)\}) > Y_1^*(v, g_2)$ and $Y_3^*(v, g_3) = 0.8 > Y_3^*(v, \{(13)\}) = 0.4$.

Finally, consider the 2-person link formation game T with player set $\{1, 3\}$ generated from the original game by fixing the strategies of players 2 and 4 at $s_2 = \{1\}$, $s_4 = \{3\}$. Routine inspection yields that there is *no* Nash equilibrium in this game. This shows that g is weakly stable.

In order to rule out inefficient graphs from being stable, we need to give some coalition the ability to deviate credibly. However, the allocation rule constructed earlier fails to ensure this essentially because agents can severe links and become the "residual claimant" in the new graph. For instance, notice that in the previous example, if 3 "deviates" from g_1 to g_2 by breaking ties with 4, then 3 becomes the residual claimant in g_2. Similarly, given g_2, 1 breaks links with 2 to establish $\{(13)\}$, where she is the residual claimant.

To prevent this jockeying for the position of the residual claimant, one can impose the condition that on all inefficient graphs, players are punished according to a *fixed* order. Unfortunately, while this does take care of the problem mentioned above, it gives rise to a new problem. It turns out that in some cases the *efficient* graph itself may not be (strongly) stable. The following example illustrates this.

EXAMPLE 4.15. Let $N = \{1, 2, 3, 4\}$. Let g^*, the unique efficient graph be $\{(12), (23), (34), (41)\}$, let $g = \{(12), (34)\}$. Assume that $v(g^*) = 4$ and $v(\{(ij)\}) = 1.5$ for all $\{i, j\} \subset N$. The values of the other graphs are chosen so that

$$\min_{S \subseteq N} \min_{g \in G_S} \frac{v(g)}{(|N| - 2)\,|S|} = 0.25.$$

Choose $\varepsilon = 0.25$ and let $\succ_p$ be an ordering on N such that $1 \succ_p 2 \succ_p 3 \succ_p 4$. Applying the allocation rule specified above, it follows that

$$Y_i(v, g^*) = 1 \qquad \text{for all} \quad i \in N$$

$$Y_2(v, g) = Y_4(v, g) = 1.25$$

and

$$Y_1(v, g) = Y_3(v, g) = 0.25.$$

One easily checks that the coalition $\{2, 4\}$ can deviate from the graph g^* to induce the graph g. This deviation makes both deviators better off. The symmetry of the value function on graphs of the form $\{(ij)\}$ now implies that *no* fixed order will work here.

This explains why we need to impose a restriction on the class of value functions. We impose a restriction which ensures that for some efficient graph g^*, there is a "central" agent within each component, that is, an agent who is connected to every other agent in the component. This restriction is defined formally below.

DEFINITION 4.16. A graph g is *focussed* if for each $h \in C(g)$, there is $i_h \in N(h)$ such that $(i_h\, j) \in h$ for all $j \in N(h) \setminus \{i_h\}$.

Let $\bar{V}$ be the set or all value functions v such that

(i) $v(g)=0$ only if g is completely disconnected.

(ii) There exists $g^* \in E(v)$ such that g^* is focussed.

We now assume that the class of permissible value functions is $\bar{V}$. This is a much stronger restriction than the assumption used in the previous theorem. However, there are several interesting problems which give rise to such value functions. Indeed, the two special models discussed by Jackson and Wolinsky (the *symmetric connections* and *coauthor* models) both give rise to value functions in $\bar{V}$.

Choose some $v \in \bar{V}$, and let $g^* \in E(v)$ be focussed. Assume that $(h_1^*, ..., h_K^*)$ are the components of g^*, and let i_k be the player who is connected to all other players in $N(h_k^*)$.[13]

Let $\succ_p$ be a strict order on the player set N satisfying the following conditions:

(i) $\forall i, j \in N$, if $i \in N(h_k)$, $j \in N(h_l)$ and $k < l$, then $i \succ_p j$.

(ii) $i_k \succ_p j$ for all $j \in N(h_k^*) \backslash \{i_k\}$, $k = 1, ..., K$.

So, $\succ_p$ satisfies two properties. First, all agents in $N(h_k^*)$ are ranked above agents in $N(h_{k+1}^*)$. Second, *within* each component, the player who is connected to all other players is ranked first. Finally, choose any ε satisfying (4.6).

The allocation rule Y^* is defined by the following rules. Choose any g and $h \in C(g)$.

(Rule 1) Suppose $N(h) = N(h^*)$ for some $h^* \in C(g^*)$. Then,

$$Y_i^*(v, g) = \frac{v(h)}{n_h} \qquad \text{for all} \quad i \in N(h).$$

(Rule 2) Suppose $N(h) \subset N(h^*)$ for some $h^* \in C(g^*)$. Let j_h be the "minimal" element of $N(h)$ under the order $\succ_p$. Then, for all $i \in N(h)$,

$$Y_i^*(v, g) = \begin{cases} (n_h - 1)\varepsilon & \text{if} \quad i \neq j_h \\ v(h) - (n_h - 1)^2 \varepsilon & \text{if} \quad i = j_h. \end{cases}$$

(Rule 3) Suppose $N(h) \nsubseteq N(h^*)$ for any $h^* \in C(g^*)$. Let j_h be the "minimal" element of $N(h)$ under the order $\succ_p$. Then, for all $i \in N(h)$,

[13] If more than one such player exists, then any selection rule can be employed.

$$Y_i^*(v, g) = \begin{cases} \dfrac{\varepsilon}{2} & \text{if} \quad i \neq j_h \\ v(h) - \dfrac{(n_h - 1)\varepsilon}{2} & \text{if} \quad i = j_h. \end{cases}$$

The allocation rule has the following features. First, provided a component consists of the same set of players as some component in g^*, the value of the component is divided equally amongst all the agents in the component. Second, punishments are levied in all other cases. The punishment is more severe if players form links *across* components in g^*.

Let s^* be the strategy profile given by $s_i^* = \{j \in N \mid (ij) \in g^*\}$ for all $i \in N$, and let $C(g^*) = \{h_1^*, ..., h_K^*\}$. We first show that if agents in components $h_1^*, ..., h_k^*$ are using the strategies s_i^*, then no group of agents in h_k^* will find it profitable to deviate. Moreover, this is independent of the strategies chosen by agents in components coming "after" h_k^*.

Lemma 4.17. *Let $v \in \bar{V}$. Suppose s is the strategy profile given by $s_i = s_i^*$ $\forall i \in N(h_k^*)$, $\forall k = 1, ..., \bar{K}$ where $\bar{K} \leqslant K$. Then, there is no s' such that $f_i^\gamma(s') > f_i^\gamma(s)$ for all $i \in T$ where $s_i' = s_i^*$ for all $i \in N(h_k^*)$, $k < \bar{K}$ and $T = \{i \in N(h_{\bar{K}}^*) \mid s_i' \neq s_i\}$.*

Proof Consider the case $\bar{K} = 1$. Let g be the graph induced by s. Note that $h_1^* \in C(g)$.

Consider any s', and let g' be the graph induced by s'. Suppose $T = \{i \in N(h_1^*) \mid s_i \neq s_i'\} \neq \emptyset$.

Case (1): There exists $h \in C(g')$ such that $N(h) = N(h_1^*)$.
In this case, Rule (1) applies, and we have

$$Y_i^*(v, g') = \frac{v(h)}{|N(h)|} \leqslant \frac{v(h_1^*)}{|N(h_1^*)|} = Y_i^*(v, g) \qquad \forall i \in N(h_1^*).$$

So no $i \in N(h_1^*)$ benefits from the deviation.

Case (2): There exists $h \in C(g')$ such that $N(h) \cap N(h_1^*) \neq \phi$, and $N(h) \not\subseteq N(h_1^*)$.
In this case, Rule (3) applies, and we have

$$Y_i^*(v, g') = \frac{\varepsilon}{2} < Y_i^*(v, g) \ \forall i \in N(h_1^*) \cap N(h).$$

Noting that $N(h_1^*) \cap N(h) \cap T \neq \emptyset$, we must have $f_i^\gamma(s) > f_i^\gamma(s')$ for some $i \in T$.

Case (3): There exists $h \in C(g')$ such that $N(h) \subset N(h_1^*)$.

Noting that there is i_1 who is connected to everyone in $N(h_1^*)$, either $i_1 \in T$ or $T = N(h)$. If $i_1 \in T$, then since $f_{i_1}^{\gamma}(s') \leqslant (n_h - 1)\,\varepsilon < f_{i_1}^{\gamma}(s)$, the lemma is true. Suppose is $i_1 \notin T$. Ruling out the trivial case where a single agent breaks away,[14] we have $|T| \geqslant 2$. From Rule 2 or Rule 3, at least one of the agents must be worse off.

Hence, in all possible cases, there is some $i \in T$ who does not benefit from the deviation.

The proof can be extended in an obvious way for all values of $\bar{K}$.

LEMMA 4.18. *Let $v \in \bar{V}$. Let g be the graph induced by a strategy profile s. Suppose there exists $h \in C(g)$ such that $N(h) \subset N(h_1^*)$. Then, g is not weakly stable.*

Proof. If s is not a Nash equilibrium of $\Gamma(\gamma)$, then there is nothing to prove. So, assume that s constitutes a Nash equilibrium.

We will prove the lemma by showing that there is a credible deviation from s for a coalition $D \subset N(h_1^*)$, $|D| = 2$. The game induced on the coalition D is defined as $\Gamma(\gamma, s_{N \setminus D}) = \langle D, \{S_i\}_{i \in D}, \bar{f}^{\gamma} \rangle$ where $\bar{f}_j^{\gamma}(s'_D) = Y_j^*(v, g(s'_D, s_{N \setminus D}))$ for all $j \in D$. We show that there is a Nash equilibrium in this two-person game which Pareto-dominates the payoff corresponding to s.

Suppose there is $i \in N(h_1^*) \setminus N(h)$, $j \notin N(h_1^*)$ such that $(ij) \in g$. Then, $Y_i^*(v, g) = \varepsilon/2$. Since s is a Nash equilibrium, this implies that i by a unilateral deviation cannot induce a graph g' in which i will be in some component such that $N(h') \subseteq N(h_1^*)$.

Now, let j be the $\succ_p$-maximal agent in $N(h)$. Consider the coalition $D = \{i, j\}$. Choose $s'_i = \{j\}$, and let s'_j be the best response to s'_i in the game $\Gamma(\gamma, s_{N \setminus D})$. Then, (s'_i, s'_j) must be a Nash equilibrium in $\Gamma(\gamma, s_{N \setminus D})$.[15] Using Rule (2), it is trivial to check that both i and j gain by deviating to s' from s.

Hence, we can now assume that if $N(h) \subset N(h_1^*)$, then there exist $\{h_1, ..., h_L\} \subseteq C(g)$ such that $N(h_1^*) = \bigcup_{i=1,...,L} N(h_i)$.[16] Note that $L \geqslant 2$.

W.l.o.g., let 1 be the $\succ_p$-maximal agent in $N(h_1^*)$, and $1 \in N(h_1)$. Let i be the $\succ_p$-maximal agent in $N(h_2)$, and let $D = \{1, i\}$.

Suppose $L > 2$. Then, consider $\bar{s}_1 = s_1 \cup \{i\}$, and let $\bar{s}_i$ be the best response to $\bar{s}_1$ in the game $\Gamma(\gamma, s_{N \setminus D})$. Note that 1 can never be the residual claimant in any component, and that $1 \in \bar{s}_i$. It then follows that $(\bar{s}_1, \bar{s}_i)$ is

[14] The agent then gets 0.

[15] The fact that i has no profitable deviation from s'_i follows from the assumption that the original strategy profile is a Nash equilibrium.

[16] Again, we are ignoring the possible existence of isolated individuals.

a Nash equilibrium in $\Gamma(\gamma, s_{N\setminus D})$ which Pareto-dominates the payoffs (of D) corresponding to the original strategy profile s.

Suppose $L=2$. Let $\bar{S}=\{\bar{s} \mid \bar{s}=(\bar{s}_1, \bar{s}_i, \bar{s}_{-D})$ for some $(\bar{s}_1, \bar{s}_i) \in S_1 \times S_i$. Let $\bar{G}$ be the set of graphs which can be induced by D subject to the restriction that both 1 and i belong to a component which is connected on $N(h_1^*)$. Let $\bar{g}$ be such that $v(\bar{g})=\max_{g \in \bar{G}} v(g)$, and suppose that $\bar{s}$ induces $\bar{g}$. Then, note that $i \in \bar{s}_1$ and $1 \in \bar{s}_i$.

Now, $Y_1^*(v, \bar{g})=Y_i^*(v, \bar{g})=v(\bar{g})/n_h$. Clearly, $Y_j^*(v, \bar{g})>Y_j^*(v, g)$ for $j \in D$. If $(\bar{s}_1, \bar{s}_i)$ is a Nash equilibrium in $\Gamma(\gamma, s_{N\setminus D})$, then this completes the proof of the lemma. Suppose $(\bar{s}_1, \bar{s}_i)$ is not a Nash equilibrium of $\Gamma(\gamma, s_{N\setminus D})$. Then, the only possibility is that i has a profitable deviation since 1 can never become the residual claimant. Let $\tilde{s}_i$ be the best response to $\bar{s}_1$ in $\Gamma(\gamma, s_{N\setminus D})$. Note that $1 \in \tilde{s}_i$. Let $\tilde{g}$ denote the induced graph. We must therefore have $Y_1^*(v, \tilde{g})>Y_1^*(v, g)$.[17] Obviously, $Y_i^*(v, \tilde{g})> Y_i^*(v, g)$. Since $\bar{s}_1$ is also a best response to $\tilde{s}_i$ in $\Gamma(\gamma, s_{N\setminus D})$, this completes the proof of the lemma.

We can now prove the following.

THEOREM 4.19. *Let $v \in \bar{V}$. Then, there exists a component balanced allocation rule Y satisfying the following*

(i) *The set of strongly stable graphs is nonempty.*

(ii) *If g is weakly stable, then $g \in E(v)$.*

(iii) *Y is anonymous over the set of weakly stable graphs.*

Proof. Clearly, the allocation rule Y defined above is component balanced. We first show that the efficient graph g^* is strongly stable by showing that s^* is a strong Nash equilibrium of $\Gamma(\gamma)$.

Let $C(g^*)=\{h_1^*, ..., h_K^*\}$.

Let $s \neq s^*$, g be the graph induced by s, and $T=\{i \in N \mid s_i \neq s_i^*\}$. Let $t^*=\arg\min_{1 \leqslant j \leqslant K} s_i \neq s_i^*$ for some $i \in N(h_j^*)$.

By Lemma 4.17, it follows that at least one member in $N(h_{t^*}) \cap T$ does not profit by deviating from the strategy s^*. This shows that the graph g^* is strongly stable.

We now show that if g is not efficient, then it cannot be weakly stable. Let s be a strategy profile which induces the graph g. We have the following cases.

Case (1a): There exists $h \in C(g)$ such that $N(h_1^*)=N(h)$ and $v(h)<v(h_1^*)$.

Suppose all individuals i in $N(h_1^*)$ deviate to s_i^*. Clearly, all individuals in $N(h_1^*)$ gain from this deviation. Moreover Lemma 4.17 shows that no

[17] This follows since 1 is now in a component containing more agents.

subcoalition of $N(h_1^*)$ has any further profitable deviation. Hence, s cannot be a CPNE of $\Gamma(\gamma)$ in this case.

Case (1b): There does not exist $h \in C(g)$ such that $N(h) \subseteq N(h_1^*)$.

In this case all players in $N(h_1^*)$ are either isolated (in which case they get zero) or they are in (possibly different) components which contain players not in $N(h_1^*)$. Using Rule (3) of the allocation rule, it follows that

$$Y_i(v, g) \leqslant \frac{\varepsilon}{2} \qquad \forall i \in N(h_1^*).$$

So all players in $N(h_1^*)$ can deviate to the strategy s_i^*. Obviously, this will make them strictly better off. That this is a credible deviation follows from Lemma 4.17.

Case (1c): There exists $h \in C(g)$, such that $N(h) \subset N(h_1^*)$.

In this case, it follows from Lemma 4.18 that there is a credible deviation for a coalition $D \subset N(h_1^*)$.

Case (2): If there exists $h \in C(g)$ such that $N(h) = N(h_1^*)$ and $v(h) = v(h_1^*)$, then apply the arguments of Case 1 to h_2^* and so on.

The preceding arguments show that if g is weakly stable, then:

(i) $N(h_i) = N(h_i^*)$ for each $i \in \{1, ..., K\}$.

(ii) $v(h_i) = v(h_i^*)$ for each $i \in \{1, ..., K\}$.

These show that all weakly stable graphs must be efficient. Furthermore, it follows from Rule (1) that Y is anonymous on all such graphs. This completes the proof of the theorem. ∎

Notice that in both Theorems 4.13 and 4.19, we have imposed the requirement that the allocation rule satisfy component balance on *all* graphs, and not just on the set of stable (or weakly stable) graphs. This raises the obvious question as to why the two properties of component balance and anonymity have been treated asymmetrically in the paper.

The answer lies in the fact that component balance has a *strategic* role, while anonymity is a purely ethical property. Consider, for instance, the "equal division" allocation rule which specifies that each agent gets $v(g)/n$ on all graphs g. This rule violates component balance.[18] Let the value function be such that $(v(\{12\})/2) > (v(g^*)/n)$ where g^* is some efficient graph. Then, given the equal division rule, agents i and j both do strictly better by breaking away from the rest of the society since the total reward given to them by this allocation rule is less than what they can get by

[18] The referee rightly points out that this rule implements the set of efficient graphs.

themselves. On the other hand, Theorems 4.13 and 4.19 show that *some* allocation rules which are component balanced ensure that no set of agents wants to break away.

Readers will notice the obvious analogy with the literature on implementation. There, mechanisms which waste resources "out of equilibrium" will not be renegotiation-proof since all agents can move away to a Pareto-superior outcome. Here, the violation of component balance implies that *all* agents in some *component* can agree on a jointly better outcome.

There is also another logical motivation which can be provided for this asymmetric treatment of component balance and anonymity.[19] In view of the Jackson–Wolinsky result, one or both the conditions must be relaxed in order to resolve the tension between stability and efficiency. This paper shows that simply relaxing anonymity out of equilibrium is sufficient. Since we have also argued that the violation of ethical conditions such as anonymity on graphs which are not likely to be observed is not a matter for concern, our results suggest an interesting avenue for avoiding the potential conflict between stability and efficiency in the context of this framework.

5. CONCLUSION

The central theme in this paper has been to examine the possibility of constructing allocation rules which will ensure that efficient networks of agents form when the individual agents decide to form or severe links amongst themselves. Exploiting the insights provided by Jackson and Wolinsky [8], it is shown that in general it may not be possible to reconcile efficiency with stability if the allocation rule is required to be anonymous on *all* graphs.

However, we go on to argue that if our prediction is that *only* efficient graphs will form, then the requirement that the allocation rule be anonymous on all graphs is unnecessarily stringent. We suggest that a "mechanism design" approach is more appropriate and show that under almost all value functions, the nonempty set of (strongly) stable graphs will be a subset of the efficient graphs under an allocation rule which is anonymous on the domain of strongly stable graphs. A stronger domain restriction allows us to prove that the above result also holds when strong stability is replaced by weak stability. Since these allocation rules will treat agents symmetrically on the graphs which are "likely to be observed," it seems that stability can be reconciled with efficiency after all.

[19] We are grateful to the Associate Editor for this suggestion.

REFERENCES

1. R. Aumann and R. Myerson, Endogenous formation of links between players and coalitions: An application of the Shapley value, *in* "The Shapley Value" (A. Roth, Ed.), Cambridge Univ. Press, Cambridge, UK, 1988.
2. B. D. Bernheim, B. Peleg, and M. Whinston, Coalition-proof Nash equilibria I. Concepts, *J. Econ. Theory* **42** (1987), 1–12.
3. B. Dutta, A. van den Nouweland, and S. Tijs, "Link Formation in Cooperative Situations," Discussion Paper 95-02, Indian Statistical Institute, New Delhi, 1995.
4. S. Goyal, "Sustainable Communication Networks," Tinbergen Institute Discussion Paper TI 93-250, 1993.
5. S. Hart and M. Kurz, Endogenous formation of cooperative structures, *Econometrica* **51** (1983), 1047–1064.
6. K. Hendricks, M. Piccione, and G. Tan, The economics of hubs: The case of monopoly, *Rev. Econ. Stud.* **62** (1995), 83–100.
7. D. Henriet and H. Moulin, Traffic-based cost allocation in a network, *Rand J. Econ.* **27** (1995), 332–345.
8. M. Jackson and A. Wolinsky, A strategic model of economic and social networks, *J. Econ. Theory* **71** (1996), 44–74.
9. J. Landa, The enigma of the Kula ring: Gift exchanges and primitive law and order, *Int. Rev. Law Econ.* **3** (1983), 137–160.
10. T. Marschak and S. Reichelstein, Communication requirements for individual agents in network mechanisms and hierarchies, *in* "The Economics of Information Decentralization: Complexity, Efficiency and Stability" (J. Ledyard, Ed.), Kluwer Academic Press, Boston, 1993.
11. R. Myerson, Graphs and cooperation in games, *Math. Oper. Res.* **2** (1977), 225–229.
12. R. Myerson, "Game Theory: Analysis of Conflict," Harvard Univ. Press, Cambridge, 1991.
13. A. van den Nouweland, "Games and Graphs in Economic Situations," Ph.D. thesis, Tilburg University, The Netherlands, 1993.
14. C. Qin, Endogenous formation of cooperative structures, *J. Econ. Theory* **69** (1996), 218–226.
15. W. Sharkey, Network models in economics, *in* "The Handbook of Operations Research and Management Science," 1993.
16. B. Wellman and S. Berkowitz, "Social Structure: A Network Approach," Cambridge Univ. Press, Cambridge, UK, 1988.