Network Topology and the Efficiency of Equilibrium

IGAL MILCHTAICH*

Department of Economics, Bar-Ilan University,

52900 Ramat-Gan, ISRAEL

First draft

June 25, 2001

Abstract. Different kinds of networks, such as transportation, communication, computer, and supply networks, are susceptible to similar kinds of inefficiencies. These arise when congestion externalities make each user's cost depend on the other users' choices of routes. If each user chooses the least expensive (e.g., fastest) route from the users' common point of origin to the common destination, the result may be inefficient in the sense that there is an alternative assignment of routes to users that reduces the costs of all users. However, this may happen only for certain kinds of network topologies. This paper gives several alternative characterizations of networks in which inefficiencies may occur. In particular, a necessary and sufficient condition for inefficiency is that one of several specific, simple networks is embedded in the network.

Keywords: Congestion, network topology, Braess's paradox, transportation networks, Wardrop equilibrium.

* E-mail: milchti@mail.biu.ac.il. Home page: http://faculty.biu.ac.il/~milchti.

1. INTRODUCTION

In transportation and other kinds of networks, congestion externalities are a potential source of inefficiency. A remarkable example of this, known as Braess's paradox (see, e.g., Nagurney, 1999), is shown in Figure 1. Cars arrive at a constant rate at vertex *o* of the depicted network and leave it at vertex *d*. The network consists of three fast roads (e_1, e_4, and e_5) and two slow detours (e_2 and e_3). The travel time on each road is an increasing function of the flow through it, or the average number of vehicles passing a fixed point in the road per unit of time. (This is a reasonable assumption if the density of vehicles on the road is relatively low, so that the flow is well below the road's capacity. See Sheffi, 1985, Chapter 13, and Figure 1.8.) However, regardless of the flow, the travel time on the route consisting of the three fast roads is shorter than on any of the alternative routes. Therefore, at equilibrium, all vehicles use that route. The travel time on the network is then 21 minutes. Suppose, however, that the transverse road, e_5, is closed, or its condition declines so that the travel time on it becomes similar to the travel time on each of the two detours. The new cost curve is higher than the old one: the travel time corresponding to every flow through e_5 is longer than before. As a result of the change in costs, the old equilibrium is replaced by a new one, in which the transverse road is not used: half the vehicles go through the left route (e_1 and e_3), and half through the right route (e_2 and e_4). Paradoxically, the new travel time is shorter than before, 20 minutes. The reason for this is that the motorists' choice of routes is only guided by concern for their own good; it does not take others' welfare into consideration. This selfish attitude results in an overuse of the fast roads, and consequently an inefficient equilibrium.

Braess's paradox is not limited to transportation networks only. There is by now a moderately large literature showing that this or similar paradoxes may also occur in such diverse networks as computer and telecommunication networks, electric circuits, and mechanical systems. Remarkably, much of this literature (e.g., Frank, 1981; Cohen and Horowitz, 1991; Cohen and Jeffries, 1997) is concerned with the same network shown in Figure 1, the Wheatstone network. As it turns out, there is a good reason for this. This paper shows that it is essentially the <u>only</u> kind of network in which Braess's paradox can occur. More precisely, a necessary and sufficient condition for the existence of <u>some</u> cost function for which the paradox occurs is that the network has an embedded Wheatstone network. In networks without this property, so-called series-parallel networks, Braess's paradox cannot occur. Several alternative characterizations of series-parallel networks are given below.

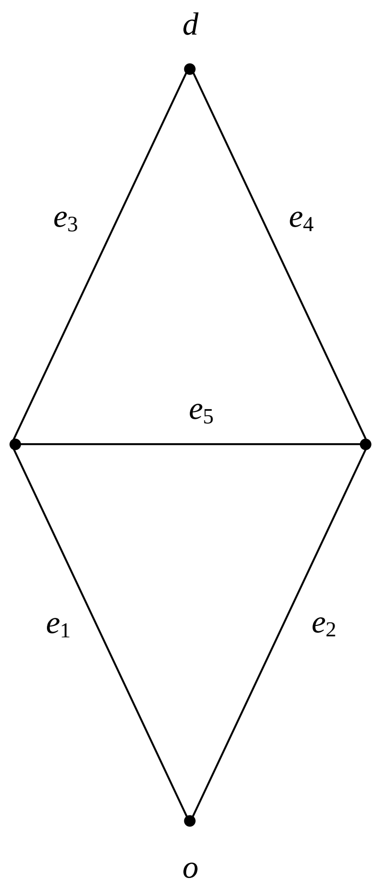
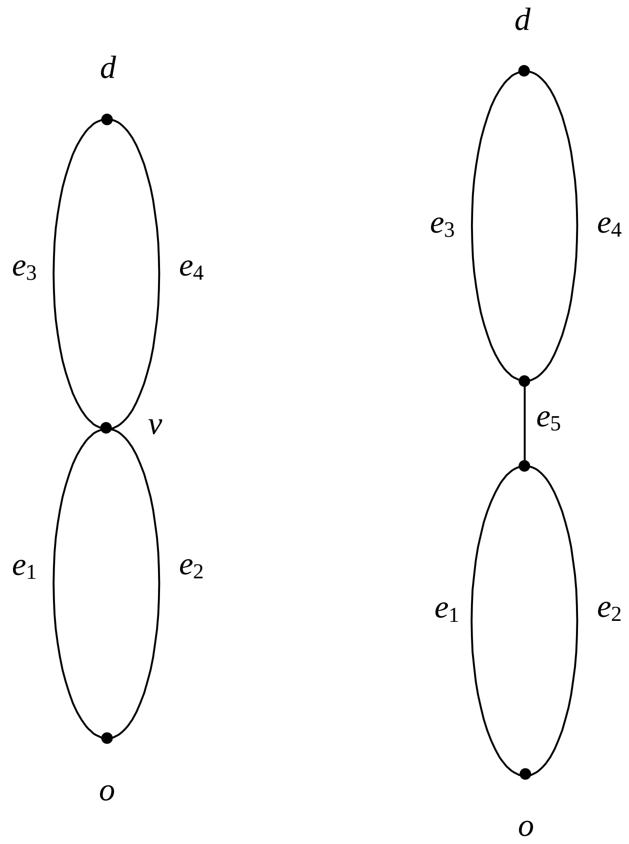


Figure 1. Braess's paradox. The travel time on each edge is an increasing function of the fraction x of the total flow from o to d passing through the edge. The travel times, in minutes, are $1 + 6x$ for e_1 and e_4, and $15 + 2x$ for e_2 and e_3. If the travel time on e_5 is also given by $1 + 6x$, then, at equilibrium, the entire flow from o to d passes through that edge. The total travel time from o to d is then 21 minutes. If, however, the travel time on e_5 is longer, and given by $15 + 2x$, then there is no flow through that edge. The equilibrium travel time is then <u>shorter</u>, 20 minutes.

The emphasis in this paper is on network topologies for which some cost functions giving rise to inefficiencies exist. Other papers, in contrast, put the emphasis on the cost functions themselves. For example, Steinberg and Zangwill (1983) and Dafermos and Nagurney (1984) derive formulas yielding, under certain conditions, the change in users' costs induced by the creation of additional routes. These formulas can, in principle, be used to determine whether a form of the Braess's paradox occurs in the network. They are, however, rather complicated. Calvert and Keady (1993) consider the total power loss in a network in which the potential drop across the two endpoints of each edge is an increasing function of the quantity obtained by dividing the edge flow by some edge-specific conductivity factor. They show that if the functional relation between this quantity and the potential drop is given by a power function, which is the same for all edges, then, when one or more of the conductivity factors is increased, with the total flow through the network kept constant, the total power loss either decreases or remains the same. Conversely, if the functions are not all equal to some power function, and there are at least six edges, it is possible to arrange the edges so that if one of the conductivity factors is increased, the total power loss also increases. Thus, a version of Braess's paradox occurs in the network. Calvert and Keady (1993) also give a topological result (Theorem 11), which says that, in a series-parallel network, this phenomenon cannot occur. In such a network, the total power loss can only decrease or remain the same when any of the conductivity factors increases.

Braess's paradox is not the only kind of inefficiency caused by congestion externalities. Consider, for example, the series-parallel network in Figure 2(a), which represents the alternatives faced by weekend visitors to a certain seaside town, where the only attractions are the two nearby beaches. The two edges joining o and v represent the alternatives of going to the North Beach (e_1) or the South Beach (e_2) on Saturday. The two edges joining v and d represent the same two alternatives on Sunday. The South Beach is more remote, and so the cost of getting there is 2 units greater than for the North Beach. On the other hand, it is a longer beach, and therefore does not get crowded as fast. However, the additional pleasure of spending the day on an uncrowded beach never exceeds the difference in travel costs. Therefore, at equilibrium all the visitors go to the North Beach, both on Saturday and on Sunday. The crowding there then reduces each person's pleasure by 4 units. However, if people were taking turns in going to the South Beach, half of them going there on Saturday and the other half on Sunday, then the cost for all individuals would be lower, 3.5. Thus, this assignment represents a Pareto improvement over the equilibrium. The difference between this example and the one above is that, in the case of the Braess's paradox, Pareto improvement results from increasing the costs of certain facilities (e.g., increasing the travel time on the transverse road in Figure 1), thereby creating a new equilibrium that is better for everyone. In contrast, in the present example it is not possible to make everybody better off simply by increasing the costs (e.g., charging congestion-dependent entry fees to beaches). Since the networks in Figure 2 are series-parallel, Braess's paradox cannot occur, and therefore any Pareto improvement must involve non-equilibrium behavior.



(a) (b)

Figure 2. Another kind of inefficiency caused by congestion externalities. The cost of each edge in network (a) is an increasing function of the fraction x of the total flow from o to d passing through the edge. For e_1 and e_3, the cost is given by $2x$. For e_2 and e_4, it is $2 + x$. At equilibrium, only e_1 and e_3 are used, and the equilibrium cost is 4. However, this outcome is inefficient. Splitting the flow, so that half of it goes through e_1 and e_4 and half through e_2 and e_3, would reduce the cost to 3.5. A similar phenomenon occurs in network (b). Indeed, since all the routes from o to d pass through the middle edge e_5, the cost of this edge is immaterial.

The main result of this paper is that the three graphs in Figures 1 and 2 essentially represent the only kinds of network topologies in which inefficient equilibria are possible. For example, inefficiencies never arise in a network like that in Figure 3. The crucial difference between this and the other networks mentioned above is that routes in this network are independent in the sense that each route contains at least one edge that is not part of any other route. The first paper to show a connection between the independence of the routes in a network (or, rather, a property equivalent to it) and the efficiency of the equilibria is Holzman and Law-Yone (1997). This paper considers, in fact, a larger class of strategy spaces, of which routes in networks constitute a sub-class. An explicit treatment of transportation networks can be found in Law-Yone (1995). For networks, the main result of Holzman and Law-Yone (1997) can be stated as follows: If the network has independent routes, then for every cost function all equilibria are weakly Pareto efficient and, moreover, are strong in the sense that no group of users can make all its members better off by changing their choices of routes. Conversely, if routes are not independent, then there is a cost function for which none of the equilibria is even weakly Pareto efficient. The main difference between Holzman and Law-Yone (1997) and the present paper is that these authors consider games with a finite number of players, each of whom has a non-negligible effect on the others. The present paper, in contrast, assumes there is a continuum of users. This may be viewed as a mathematical idealization of a very large population of individuals, each with a nearly negligible effect on the others. There are several substantial differences between these two cases. One of them is that, in the latter but not in the former case, the connection between the independence of the routes and the efficiency of the equilibria also holds for heterogeneous populations, in which not all users have the same cost function.

Users may differ in the innate quality they assign to the various alternatives, or in the degree by which they are affected by congestion. For example, some motorists may be concerned primarily with the travel time, and others with the distance traveled. In small populations, heterogeneity is a potential source of inefficiency. This can be demonstrated by the simple two-user, two-facility example in which each user has a different favorite facility, but would rather not use it than share it with the other user (see Milchtaich, 1996). In this example, there are two (pure-strategy) Nash equilibria, one of which is worse for both users. In contrast, if there is a continuum of users, then one Nash equilibrium may strictly Pareto dominate another only if one of the networks in Figures 1 and 2 is embedded in the network. In a network that does not have this property, i.e., one with independent routes, all Nash equilibria are efficient—for heterogeneous as well as homogeneous populations.

An intermediate model between that of a finite population of users and that of a continuum of users is the one in which flow is continuous but the population of users is finite (see, for example, Orda et al., 1993). Each user controls a fixed portion of the flow and distributes it so as to minimize the user's total cost. While the results of the present paper are not directly applicable, it is likely that connections similar to those established here between network topology and the efficiency of the equilibria also hold for this model. Yet another line of investigation concerns the social optimality of the equilibria, i.e., whether, at equilibrium, the total cost is minimized—a condition more demanding than Pareto efficiency. This may not be answered by only considering the network topology; it is also necessary to know the functional form of the cost functions. For networks that consist of several edges connected in parallel, conditions for social optimality of the equilibria are established in Milchtaich (2001).

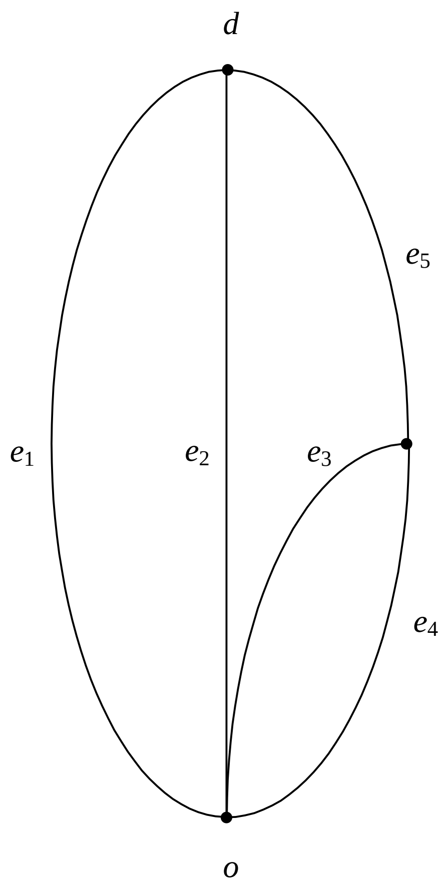


Figure 3. A network with independent routes.

2. GRAPH THEORETIC PRELIMINARIES

2.1. Basic terminology

An undirected multigraph consists of a finite set V of vertices together with a finite set E of edges. Each edge e is associated with an unordered pair $\{u, v\}$ of distinct vertices, which the edge is said to join. These are called the end vertices of e. Thus, loops are not allowed, but more than one edge can join two vertices. An edge e and a vertex v are said to be incident with each other if v is an end vertex of e. A walk of length n ($n \geq 0$) is an alternating sequence s of vertices and edges v_0, e_1, v_1, …, v_{n-1}, e_n, v_n, beginning and ending with vertices, in which each edge is incident with the two vertices immediately preceding and following it. The vertices v_0 and v_n are called the initial and terminal vertices of s, respectively. The walk v_n, e_n, v_{n-1}, …, v_1, e_1, v_0, which includes the same vertices and edges as s but passes them in reverse order, is denoted $-s$. If t is a walk of the form v_n, e_{n+1}, v_{n+1}, …, v_{m-1}, e_m, v_m, the initial vertex of which is the same as the terminal vertex of s, then v_0, e_1, …, e_n, v_n, e_{n+1}, …, e_m, v_m is also a walk, denoted $s + t$. A section of s is any walk of the form v_{n_1}, e_{n_1+1}, v_{n_1+1}, …, v_{n_2-1}, e_{n_2}, v_{n_2}, with $0 \leq n_1 \leq n_2 \leq n$. If all the vertices (and, hence, all the edges) in a walk s are distinct, then each section of s is uniquely identified by its initial vertex u and terminal vertex v, and may therefore be denoted by s_{uv}. If the section is of length zero, i.e., has no edges, then u and v coincide. If it is of length one, i.e., has a single edge, than u and v are the end vertices of that edge. In this case, the section may be viewed as an indication of the direction in which s passes through the edge.

2.2. Two-terminal networks

A two-terminal network (network, for short) is an undirected multigraph together with a distinguished ordered pair of distinct vertices, o (for "origin") and d (for "destination"), such that each vertex and each edge belong to at least one walk in which the initial vertex is o, the terminal vertex is d, and all the vertices are distinct. Any walk with these properties will be called a route. The set of all routes in a network is denoted $\mathcal{R}$.

Two networks G and G' will be said to be isomorphic if there is a one-to-one correspondence between their vertices and between their edges such that (i) the incidence relation is preserved and (ii) the origin and destination in G are paired with the origin and destination in G', respectively. A network G is embedded in a network G' if G' is isomorphic to a network obtained from G by carrying out any number of times one or more of the following operations (see Figure 4):

1. Subdividing an edge; i.e., replacing it with two edges with a single common end vertex.
2. Adding an edge that joins two existing vertices.
3. "Extending" the origin or the destination; i.e., adding an edge joining *o* or *d* and another, new vertex, which becomes the new origin or destination, respectively.

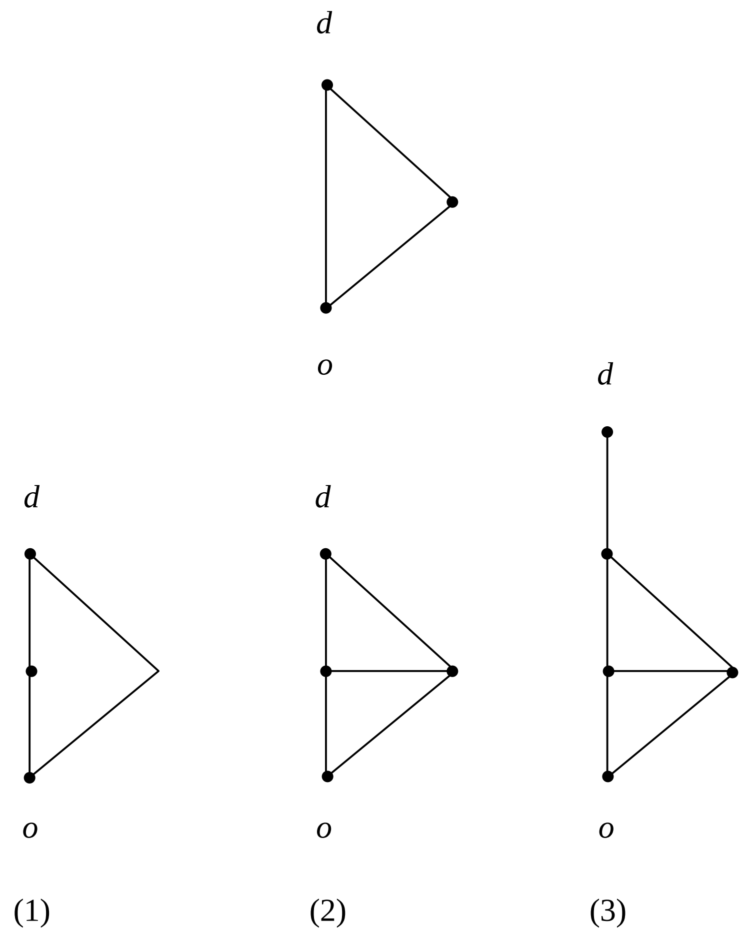


Figure 4. The upper network is embedded in each of the three lower ones. These are obtained from it by carrying out in sequence the following operations: (1) subdividing an edge, (2) adding an edge that joins two existing vertices, and (3) extending the destination.

Two networks G' and G'' with the same origin–destination pair, but no other common vertices or edges, may be connected in parallel. The set of vertices in the resulting network G is the union of the sets of vertices in G' and G'', and similarly for the set of edges. The origin and destination in G are the same as in G' and G''. Two networks G' and G'' with a single common vertex (and, hence, without common edges), which is the destination in G' and the origin in G'', may be connected in series. The set of vertices in the resulting network G is the union of the sets of vertices in G' and G'', and similarly for the set of edges. The origin in G coincides with the origin in G', and the destination is the destination in G''.

A network is said to be series-parallel if no two routes pass through any edge in opposite directions. The two networks in Figure 2 are series-parallel. The Wheatstone network in Figure 1 is not series-parallel, since there are two routes passing through e_5 in opposite directions. In fact, as the following proposition shows, the Wheatstone network is part of any series-parallel network. This result is very similar to one of Duffin (1965, Theorem 1).

Proposition 1. A network G is series-parallel if and only if the network in Figure 1 is not embedded in it.

As noted by Riordan and Shannon (1942), series-parallel networks can also be defined recursively: A network is series-parallel if and only if it can be constructed from single edges by carrying out any number of times the operations of connecting networks in series or in parallel. Hence the term “series-parallel.” The following proposition establishes this.

Proposition 2. A network G is series-parallel if and only if

(i) it has a single edge only; or

(ii) it is the result of connecting two series-parallel networks in parallel; or

(iii) it is the result of connecting two series-parallel networks in series.

One corollary of Proposition 2 is that every series-parallel network is planar and, moreover, remains so when a new edge, joining o and d, is added to it. Equivalently, every series-parallel network can be embedded in the plane in such a way that o and d lie on the exterior face, or boundary. Using Proposition 2, this result can easily be proved by induction on the number of edges. A practical way of verifying that a given network is series-parallel is suggested by the third condition in the following proposition.

Proposition 3. For every network *G*, the following three conditions are equivalent:

(i) *G* is series-parallel.

(ii) For every pair of distinct vertices *u* and *v*, if *u* precedes *v* in some route *r* containing both vertices, then *u* precedes *v* in all such routes.

(iii) The vertices can be indexed in such a way that, along each route, they have increasing indices.

A network with independent routes is one in which every route contains at least one edge that does not belong to any other route. An example of a network with independent routes is shown in Figure 3. As shown below, such a network is necessarily series-parallel. The converse, however, is false. For example, the two networks in Figure 2 are series-parallel but the routes in them are not independent. In fact, it follows from the next proposition that these are essentially the only such networks.

Proposition 4. A network *G* is a network with independent routes if and only if none of the networks in Figures 1 and 2 is embedded in it.

The result that every network with independent routes is series-parallel follows as an immediate corollary from Propositions 1 and 4. It can also be deduced from the following recursive characterization of such networks, due to Law-Yone (1995). This characterization differs from the one for series-parallel networks (Proposition 2) only in condition (iii), which is stronger than the corresponding condition there.

Proposition 5. A network *G* is a network with independent routes if and only if

(i) it has a single edge only; or

(ii) it is the result of connecting in parallel two networks with independent routes; or

(iii) it is the result of connecting in series a network with independent routes and a network with a single edge.

In every network, each route r has a unique set of edges, and may therefore be identified with a unique binary vector, coordinate e of which is 1 if edge e belongs to r and 0 otherwise. This vector can be viewed as an element of the vector space $\mathbb{Z}_2^{|E|}$, where $|E|$ is the number of edges in the network and $\mathbb{Z}_2$ is the field of the integers modulo 2. A set of routes will be said to be independent if the corresponding set of vectors is linearly independent in $\mathbb{Z}_2^{|E|}$. Equivalently, a set of routes is independent if it is not possible to write one of the corresponding vectors as the (component-wise) sum modulo 2 of some of the others. As the following proposition shows, a network with independent routes is characterized by the property that the set of all routes is independent.[1] Hence the term. An equivalent property is that the set of all routes does not contain a "bad configuration" (Holzman and Law-Yone, 1997). A bad configuration is a triplet of routes such that the first route contains some edge e_1 that does not belong to the second route, the second route contains some edge e_2 that does not belong to the first route, and the third route contains both e_1 and e_2. Another property that characterizes networks with independent routes is that pairs of routes never merge only in their middle; in other words, any common section must extend either to o or to d.

Proposition 6. For every network G, the following four conditions are equivalent:

(i) The set $\mathcal{R}$ of all routes in G is independent.

(ii) A triplet of routes constituting a bad configuration does not exist.

(iii) For every pair of distinct routes r and s and every vertex v common to both routes, either the section r_{ov} (which consists of v and all the vertices and edges preceding it in r) is equal to s_{ov}, or r_{vd} is equal to s_{vd}.

(iv) G is a network with independent routes.

[1] Note that independence is defined with respect to $\mathbb{Z}_2$, not (the real field) $\mathbb{R}$. For example, the network in Figure 1 is not a network with independent routes, despite the fact that the vectors representing its four routes are linearly independent in $\mathbb{R}^{|E|}$. This is because, in $\mathbb{Z}_2^{|E|}$, each of these four vectors is equal to the sum of the others.

3. FLOWS AND COSTS

A flow in a network is a specification of a nonnegative route flow f_r for each route r. It can be written in a form of a flow vector **f**, coordinate r of which is f_r. Given the flow vector, the flow f_s through any walk s is defined as the total flow in all the routes of which s is a section:

$$f_s = \sum_{\substack{r \in \mathcal{R} \\ s \text{ is a section of } r}} f_r. \tag{1}$$

If s is a walk of length zero, consisting of a single vertex, then f_s represents the total flow in all the routes passing through that vertex. In particular, the flow through the origin,

$$f_o = \sum_{r \in \mathcal{R}} f_r,$$

represents the total origin–destination flow. It is equal, of course, to f_d, the flow through the destination. If s is a walk of length one, consisting of a single edge and its two end vertices, then f_s represents the total flow through that edge in a the direction indicated by s; it will be referred to as an edge flow. Note that each edge is associated with a pair of edge flows, one in each direction. However, in a series-parallel network, in which all routes pass through an edge in the same direction, only one of these can ever be positive. In a network with independent routes, the edge flows uniquely determine the flow vector.

It should be emphasized that, in this paper, flow is always assumed to originate in a single vertex, o, and terminate in a single vertex, d. Multiple origin–destination pairs are not allowed. This restriction can be partially circumvented by connecting all sources to a single, artificial vertex, from which all flow is assumed to originate, and similarly for the sinks. However, such a construction substantially alters the network topology.

A cost function is a vector-valued function **c** specifying the cost $c_s(\mathbf{f})$ of each walk s as a function of the flow vector **f**.[2] This cost is assumed to satisfy the following

[2] Note that all walks are assigned costs, not just routes. The costs are not assumed to be nonnegative. However, they may be thought of as such. Indeed, the assumption that costs cannot be negative is implicit in the definition of equilibrium (in the next section), which only considers routes, i.e., walks from the origin to the destination that do not pass through any vertex more than once.

montonicity condition: For every pair of flow vectors $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$, if $\hat{f}_t \geq \tilde{f}_t$ and $\hat{f}_{-t} \geq \tilde{f}_{-t}$ for all sections t of s, then $c_s(\hat{\mathbf{f}}) \geq c_s(\tilde{\mathbf{f}})$.[3] This implies, in particular, that the cost of a walk only depends on the flow through each of its sections and the flows in the opposite directions. In general, the cost may remain constant even if these flows increase. A cost function **c** will be said to be increasing if it satisfies the following condition: For every route r and every pair of flow vectors $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$, if $\hat{f}_s \geq \tilde{f}_s$ and $\hat{f}_{-s} \geq \tilde{f}_{-s}$ for all sections s of r, and there is at least one section s of length one for which $\hat{f}_s > \tilde{f}_s$, then $c_s(\hat{\mathbf{f}}) > c_s(\tilde{\mathbf{f}})$. A cost function **c** will be said to be additively separable if, for every route r, every pair of distinct vertices u and v in r, and every flow vector $\mathbf{f}$, if u precedes v in r then $c_{r_{ov}}(\mathbf{f}) = c_{r_{ou}}(\mathbf{f}) + c_{r_{uv}}(\mathbf{f})$. Additive separability implies that the cost of each route is the sum of the costs of its individual edges. Note that the cost of an edge generally depends on the direction in which it is passed through. However, in each direction, the cost is only a function of the flows through the edge in that direction and the opposite direction and (possibly) the flows through the end vertices. In a series-parallel network, where all routes pass through an edge in the same direction, the cost of passing through it in the opposite direction, as well as the effect on the cost of the flow in the opposite direction, are immaterial.

In models of transportation networks, the assumption that the cost function is additively separable is customary. Moreover, it is often assumed that there is only one direction in which each edge can be passed through (e.g., Sheffi, 1985; Bell and Iida, 1997). Thus, a two-way highway is described by a pair of edges. Correspondingly, the description of a transportation network typically involves two kinds of data: a directed graph, which describes both the physical network and the directions in which individual edges may be traveled on; and a system of associated edge cost functions, which give the travel time on each edge as a function of the edge flow. However, the model presented in Beckmann et al. (1956) differs in that it assumes all roads to be

[3] This condition is rather weak: Since it involves a potentially long list of assumptions, the set of flow vector pairs to which it applies is relatively small. Stronger, and perhaps more intuitive, monotonicity conditions could be used instead. For example, it could be required that the cost of a walk can decrease only if one or more of the relevant edge flows decrease. However, a weaker definition makes for stronger results, and it is therefore preferable from a methodological point of view. The same remark applies to the two definitions that follow.

two-way. Moreover, the cost of traveling on a road in both directions is assumed equal, and is only a function of the sum of the flows in all routes passing through the road in either direction. The model presented in the present paper subsumes both these models. In the additively separable case, each edge is associated with a pair of cost functions—one in each direction. The first transportation model described above corresponds to a case in which the cost of passing through an edge in a particular direction is prohibitively high. The second model corresponds to a case in which the two costs are equal, and only depend on the sum of the edge flows in both directions.

Dropping the additive separability assumption gains the model some generality. For example, turning restrictions can be incorporated simply by assigning very high (or infinite) costs to certain routes. The same applies to the possibility that route costs are affected by the flow through one or more of their vertices. Such flows may, for example, represent a crude measure of congestion at four-way stop junctions. However, neither the possibility of non-separable costs nor that of junction costs is essential for any of the results below.

4. EQUILIBRIUM

A flow vector **f*** is said to be an equilibrium if the entire flow from o to d passes through least cost routes. Mathematically, the equilibrium condition is:

(2) For every route r with $f_r^* > 0$,

$$c_r(\mathbf{f^*}) = \min_{s \in \mathcal{R}} c_s(\mathbf{f^*}).$$

In this case, the above minimum, denoted c^*, is the equilibrium cost. In the transportation literature, a flow vector satisfying (2) is known as a Wardrop, or user, equilibrium. This condition expresses the principle, formulated by Wardrop and others (see references in Nagurney, 1999, p. 151), that, at equilibrium, the travel time on all used routes is equal, and less than or equal to the travel time that would be experienced by of a single vehicle on any unused route (Sheffi, 1985).

The equilibrium condition (2) can also be given a variational inequality formulation, as follows (see Nagurney, 1999, Theorem 4.5):

For every flow vector **f** with the same total origin–destination flow as **f***,

$$\sum_{r \in \mathcal{R}} c_r(\mathbf{f}^*)\,(f_r^* - f_r) \le 0.$$

Standard results (e.g., Nagurney, 1999, Theorem 1.4) then imply that, if the cost function is continuous, then for every $d \geq 0$ there exists an equilibrium with a total origin–destination flow of d. In general, there can be more than one such equilibrium. However, as the following proposition shows, if the network is series-parallel and the cost function is additively separable, then the total origin–destination flow uniquely determines the equilibrium cost. Moreover, in such a setting, increasing the edge costs or the total origin–destination flow cannot result in a decreasing equilibrium cost.

Proposition 7. Let $\hat{\mathbf{c}}$ and $\tilde{\mathbf{c}}$ be two additively separable cost functions for the same series-parallel network, such that $\hat{c}_r(\mathbf{f}) \geq \tilde{c}_r(\mathbf{f})$ for all routes r and flow vectors $\mathbf{f}$. If $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$ are equilibria with respect to $\hat{\mathbf{c}}$ and $\tilde{\mathbf{c}}$, respectively, and the total origin–destination flows satisfy $\hat{f}_o \geq \tilde{f}_o$, then the respective equilibrium costs $\hat{c}$ and $\tilde{c}$ satisfy $\hat{c} \geq \tilde{c}$. Consequently, if the cost functions and the total origin–destination flows are the same, then the equilibrium costs are also the same.

The result that, in a series-parallel network with an additively separable cost function, the equilibrium cost cannot decrease when edge costs increase, holds in the case of elastic as well as inelastic demand. In the latter case, in any equilibrium the total origin–destination flow must equal the fixed demand d. Therefore, this result follows directly from Proposition 7. In the former case, the demand is determined as a nonincreasing function of the cost (Sheffi, 1985, p. 135; Bell and Iida, 1997, p. 102). That is, the total origin–destination flow may vary, but is lower in one equilibrium than in another only if the first equilibrium cost is higher than the second. In particular, for $\hat{\mathbf{c}}$ and $\tilde{\mathbf{c}}$ as in Proposition 7 and for any pair of corresponding equilibria $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$, if $\hat{f}_o < \tilde{f}_o$ then $\hat{c} \geq \tilde{c}$. On the other hand, by Proposition 7, $\hat{f}_o \geq \tilde{f}_o$ also implies $\hat{c} \geq \tilde{c}$. Therefore, the last inequality holds unconditionally. This proves that, in the case of elastic as well as inelastic demand, a unique equilibrium cost corresponds to any additively separable cost function, and this cost either increases or remains the same when edge costs are increased.

The case of elastic demand is, in fact, not that much different from the case of fixed, inelastic demand. This can be shown by making the option of "staying at home" explicit. This option, which carries a fixed cost, is optimal if and only its cost is less than that of any route in the network. Therefore, the option of "staying at home" can be represented by a single edge, joining o and d, with a flow-independent cost. By Propositions 2 and 5, adding such an edge to a series-parallel network or one with independent routes does not affect the respective property. Also, if the cost function in the original network is additively separable, then the same is true for the enlarged one. These considerations show that, for certain purposes at least, demand may be assumed inelastic. Correspondingly, it suffices to restrict attention to flow vectors with a total origin–destination flow equal to the fixed demand.

5. EFFICIENCY OF EQUILIBRIA

Braess's paradox represents an extreme form of inefficiency. It occurs in a network G if there are two additively separable cost functions $\hat{\mathbf{c}}$ and $\tilde{\mathbf{c}}$ such that $\hat{c}_r(\mathbf{f}) \geq \tilde{c}_r(\mathbf{f})$ for all routes r and flow vectors $\mathbf{f}$, but for every equilibrium $\hat{\mathbf{f}}$ with respect to $\hat{\mathbf{c}}$ with a total origin–destination flow of unity[4] and every equilibrium $\tilde{\mathbf{f}}$ with respect to $\tilde{\mathbf{c}}$ with a similar total origin–destination flow, the respective equilibrium costs $\hat{c}$ and $\tilde{c}$ satisfy $\hat{c} < \tilde{c}$. Thus, Braess's paradox occurs when higher cost curves potentially correspond to a lower equilibrium cost. As the following theorem shows, this is the case for every network that is not series-parallel. Conversely, for a network that is series-parallel, a pair of cost functions as above never exists. This result immediately implies Theorem 11 of Calvert and Keady (1993) (for the case of two-terminal networks), which is described in the Introduction.

Theorem 1. For every network G, the following conditions are equivalent:

(i) Braess's paradox occurs in G.

(ii) G is not series-parallel.

---

[4] The total origin–destination flow can always be normalized to 1. Therefore, restricting attention to equilibria satisfying this condition involves no loss of generality.

Even though series-parallel networks never exhibit Braess's paradox, they do not always have efficient equilibria. This is demonstrated by the example in Figure 2, in which the equilibrium flow can be rearranged in such a way that the costs of all routes actually used become less than the equilibrium cost. As the next theorem shows, the reason this kind of inefficiency occurs in the networks in Figure 2 is that routes in these networks are not independent.

An equilibrium **f***, with equilibrium cost c^*, will be said to be weakly Pareto efficient if there is no flow vector **f**, with the same total origin–destination flow as **f***, such that $c_r(\mathbf{f}) < c^*$ for all routes r with $f_r > 0$. It is Pareto efficient if there is no flow vector **f**, with the same total origin–destination flow as **f***, such that $c_r(\mathbf{f}) \le c^*$ for all routes r with $f_r > 0$ and there is at least one such route r for which the inequality is strict.

Theorem 2. For every network G, the following conditions are equivalent:

(i) For every cost function, all equilibria are weakly Pareto efficient.

(ii) For every increasing cost function, all equilibria are Pareto efficient.

(iii) G is a network with independent routes.

6. HETEROGENEITY

The model presented in Section 3 can be generalized by dropping the (implicit) assumption that all users of the same route incur the same costs. This requires modifying the definition of equilibrium as well. If route costs differ across users, the equilibrium condition becomes: Each route is only used by those for whom it is a least cost route. A more formal definition follows. Let the population of users be the unit interval [0, 1]. For each user i, let $\mathbf{c}^i$ be i's cost function. This specifies the cost $c_s^i(\mathbf{f})$ of each walk s for user i as a function of the flow vector **f**, and is assumed to satisfy the condition that, for every walk s and every pair of flow vectors $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$, if $\hat{f}_t \ge \tilde{f}_t$ and $\hat{f}_{-t} \ge \tilde{f}_{-t}$ for all sections t of s, then $c_s^i(\hat{\mathbf{f}}) \ge c_s^i(\tilde{\mathbf{f}})$. A strategy profile s is an assignment of a route $s(i)$ to each user i, such that, for every route r, the set of all users i with $s(i) = r$ is Lebesgue measurable; the measure of this set is the route flow f_r. Thus, each strategy profile s gives rise to a particular flow vector, with a total origin–destination flow of unity. This flow vector will be denoted by $\mathbf{f}(s)$. A strategy profile s is a Nash equilibrium if, for every user i,

$$c^i_{s(i)}(\mathbf{f}(s)) = \min_{s\in\mathcal{R}} c_s^i(\mathbf{f}(s)). \qquad (3)$$

In this case, the right-hand side of (3) gives user i's equilibrium cost. In the special case in which all users have the same cost function, this definition essentially reduces to (2).

The case of user-dependent cost functions differs from that of a single, common cost function in a number of ways. For example, the result that, for series-parallel networks and additively separable cost functions, equilibrium costs are unique (Proposition 7), is no longer true if heterogeneity is allowed. This is shown by the following example.

Example 1. Three types of users travel from o to d in the series-parallel network shown in Figure 5. Type I users are those with $0 \le i < 1/8$, type II are those with $1/8 \le i < 1/4$, and type III are those with $1/4 \le i \le 1$. The cost functions are additively separable: for each type of user, the cost of each route is the sum of the costs of its edges. The cost of edge e_j for user i depends on i's type and on the fraction x of users using edge e_j, as detailed in the following table:

	e_1	e_2	e_3	e_4	e_5
Type I	x			$5x$	x
Type II		$5x$	x		x
Type III		x		$0.5 + x$	x

(where blank cells indicate prohibitively high costs). There is one Nash equilibrium in which all type I users take the route going through e_1 and e_4, all type II users take the route going though e_2 and e_3, and all type III users use e_5. In this Nash equilibrium, everyone's cost is 3/4. However, there is another Nash equilibrium, in which all users incur a higher cost, 5/6. In this Nash equilibrium, type III users with $5/6 \le i \le 1$ take the route going though e_2 and e_4, and all the other users use e_5. Any convex combination of these two equilibria (in terms of the proportion of users of each type using each route) is also a Nash equilibrium, with an intermediate equilibrium cost. Thus, there is a continuum of Nash equilibria. These equilibria can be Pareto ranked in the sense that, in each equilibrium, everyone's cost is higher, or everyone's cost is lower, than in each of the other equilibria. Indeed, the costs in each equilibrium are the same for all users. These costs are minimal at the first Nash equilibrium described above. Moreover, it can be shown that this Nash equilibrium is Pareto efficient, in the obvious sense (see below).

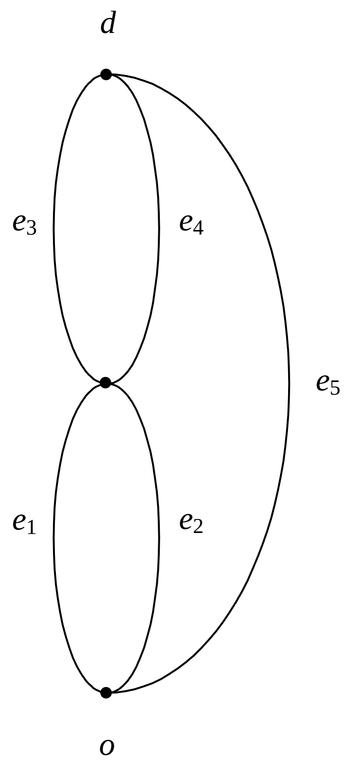


Figure 5. Population heterogeneity and inefficiency of equilibria. For details see text.

The second Nash equilibrium in Example 1 is not even weakly Pareto efficient. As in the homogeneous case (Theorem 2), this lack of efficiency may be attributed to the fact that routes in the network are not independent. As Theorem 3 below shows, in a network with independent routes Nash equilibria are always weakly Pareto efficient. And if the cost functions of all users are increasing, then all Nash equilibria satisfy a condition stronger than Pareto efficiency, which will be referred to as hyper-efficiency.

A strategy profile s is weakly Pareto efficient if there is no strategy profile t such that $c^i_{t(i)}(\mathbf{f}(t)) < c^i_{s(i)}(\mathbf{f}(s))$ for all users i. It is Pareto efficient if, for every strategy profile t,

either $c^i_{t(i)}(\mathbf{f}(t)) = c^i_{s(i)}(\mathbf{f}(s))$ for all users i,

or $c^i_{t(i)}(\mathbf{f}(t)) > c^i_{s(i)}(\mathbf{f}(s))$ for some i.

A strategy profile s will be said to be hyper-efficient if, for every strategy profile t,

(H) either $c^i_{t(i)}(\mathbf{f}(t)) = c^i_{s(i)}(\mathbf{f}(s))$ for all users i,

or $c^i_{t(i)}(\mathbf{f}(t)) > c^i_{s(i)}(\mathbf{f}(s))$ for some i with $t(i) \neq s(i)$.

In words, any effective change of route choices is harmful to some of those who change their routes. Clearly, any hyper-efficient strategy profile s is a Nash equilibrium. Indeed, it is a strong equilibrium, and even a strictly strong equilibrium.[5] This means that deviations are never profitable, not just for individuals but also for groups of users: Any deviation that makes someone in the group better off must leave someone else worse off. In a network with independent routes, and when all cost functions are increasing, the converse is also true. That is, under these conditions any Nash equilibrium is hyper-efficient and, hence, Pareto efficient and a strictly strong equilibrium.

Theorem 3. For every network G, the following conditions are equivalent:

(i) For every assignment of cost functions to users, all Nash equilibria are weakly Pareto efficient.

(ii) For every assignment of increasing cost functions to users, all Nash equilibria are hyper-efficient.

(iii) G is a network with independent routes.

While independence of the routes implies that all equilibria are efficient, even this condition does not guarantee uniqueness of the equilibrium costs if the population is heterogeneous. The following example shows this.

[5] A strategy profile s is a strictly strong equilibrium (Voorneveld at al., 1999) if, for every strategy profile t, $c^i_{t(i)}(\mathbf{f}(t)) \geq c^i_{s(i)}(\mathbf{f}(s))$ for all users i, or $c^i_{t(i)}(\mathbf{f}(t)) > c^i_{s(i)}(\mathbf{f}(s))$ for some i with $t(i) \neq s(i)$.

Example 2. Three types of users travel from *o* to *d* in the network shown in Figure 3. Type I users are those with $0 \le i < 4/13$, type II are those with $4/13 \le i < 8/13$, and type III are those with $8/13 \le i \le 1$. The cost functions are additively separable: for each type of user, the cost of each route is the sum of the costs of its edges. The cost of edge e_j for user *i* depends on *i*'s type and on the fraction *x* of users using edge e_j, as detailed in the following table:

	e_1	e_2	e_3	e_4	e_5
Type I	$3.1 + x$			$8x$	x
Type II		$8x$	$2.1 + x$		x
Type III		$0.5 + x$		x	x

(where blank cells indicate prohibitively high costs). There is one Nash equilibrium in which all type I users take the route going through e_4 and e_5, all type II users take the route going through e_3 and e_5, and all type III users use e_2. There is another Nash equilibrium, in which all type I users use e_1, all type II users use e_2, and all type III users take the route going through e_4 and e_5. For type I users, the second equilibrium cost is higher than the first. For type II and type III users, the first is higher than the second.

A yet unresolved problem is the determination of necessary and sufficient topological conditions guaranteeing that, for any assignment of increasing, additively separable cost functions to users, each user's equilibrium cost is unique. A sufficient condition for this is that the network consists of several edges connected in parallel (Milchtaich, 2000; Konishi, 2001), or is the result of connecting several such networks in series. However, this condition is not necessary. For sufficient conditions on the cost functions, which guarantee uniqueness of the equilibrium edge flows and, hence, the equilibrium costs, see, for example, Altman and Kameda (2001). The conditions are essentially these: Each edge can be passed through in one direction only; and its cost is equal to some increasing function of the edge flow, which is the same for all users, plus a user-specific constant, which does not depend on the flow.

APPENDIX

The Appendix contains the proofs of the various propositions and theorems in this paper.

Proof of Proposition 1. The three claims given below together prove the proposition and, in addition, establish the following result.

Lemma 1. A network is series-parallel if and only if it satisfies the following condition:

> For every pair of distinct vertices u and v, if u precedes v in some route r containing both vertices, then u precedes v in all such routes.

Consequently, for every two routes r and s in a series-parallel network and every vertex v common to both routes, the walk $r_{ov} + s_{vd}$ is a route (i.e., it does not pass through any vertex more than once).

CLAIM 1. Every network that satisfies the condition in Lemma 1 is series-parallel.

This is obvious. Indeed, for a network to be series-parallel it only has to satisfy the condition for all pairs of vertices u and v that are joined by some edge.

CLAIM 2. The network in Figure 1 is embedded in every network that does not satisfy the above condition.

Suppose there are two routes r and s in a network G, and two vertices u and v common to both routes, such that u precedes v in r but follows it in s. Suppose also that u and v are chosen in such a way that the length of r_{uv} is maximal. Then, any vertex u' common to r and s that precedes u in r must precede v in s, and any vertex v' common to both routes that follows v in r must follow u in s (see Figure 6). Let u' be chosen as the last vertex before u in r that also belongs to s (possibly, $u' = o$), and v' as the first vertex after v in r that also belongs to s (possibly, $v' = d$). All the edges in $r_{u'u}$, and all the vertices in this section of r with the exception of the initial and terminal ones, do not belong to s, and the same is true for $r_{vv'}$. This implies that the network in Figure 1 is embedded in the network that consists of all the vertices and edges in s, $r_{u'u}$, and $r_{vv'}$. Hence, the same is true for G.

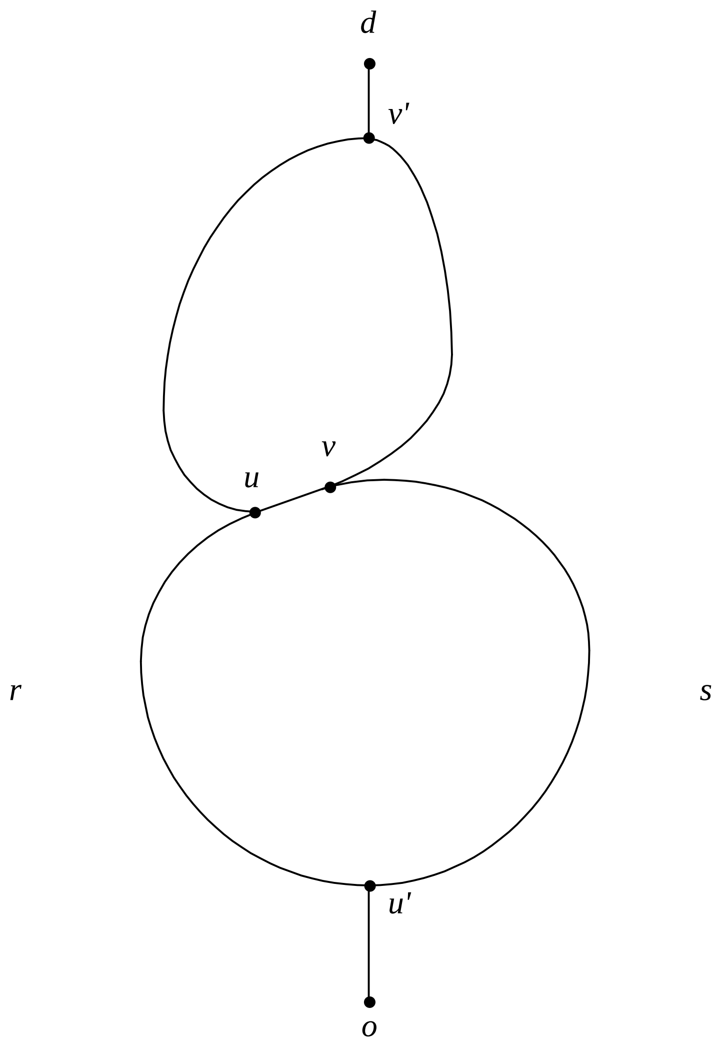


Figure 6.

CLAIM 3. The network in Figure 1 is not embedded in any series-parallel network.

This is clear: Any network with an embedded non-series-parallel network is itself not series-parallel. ■

Proof of Proposition 2. Clearly, all networks with a single edge, as well as those that are the result of connecting two series-parallel networks in series or in parallel, are themselves series-parallel. Therefore, it only has to be shown that every series-parallel network *G* with more than one edge satisfies (ii) or (iii).

CLAIM 1. In the set of all routes in *G*, the relation "route *r* and route *s* have a vertex in common, other than *o* and *d*, or the two routes are identical," is an equivalence relation.

This relation is obviously reflexive and symmetric. It remains to be shown that it is transitive. That is, if *r*, *s*, and *t* are three routes such that there is some vertex $u \neq o, d$ common to *r* and *s* and some vertex $v \neq o, d$ common to *s* and *t*, then there is also some vertex, other than *o* and *d*, common to *r* and *t*. Suppose not. Without loss of generality, it may be assumed that *u* precedes *v* in *s*, and there is no other vertex in the section s_{uv} that belongs either to *r* or to *t* (see Figure 7). This assumption implies that the network in Figure 1 is embedded in the network that consists of all the vertices and edges in *r*, *t*, and s_{uv}. However, by Proposition 1, this contradicts the assumption that *G* is series-parallel. This contradiction proves Claim 1.

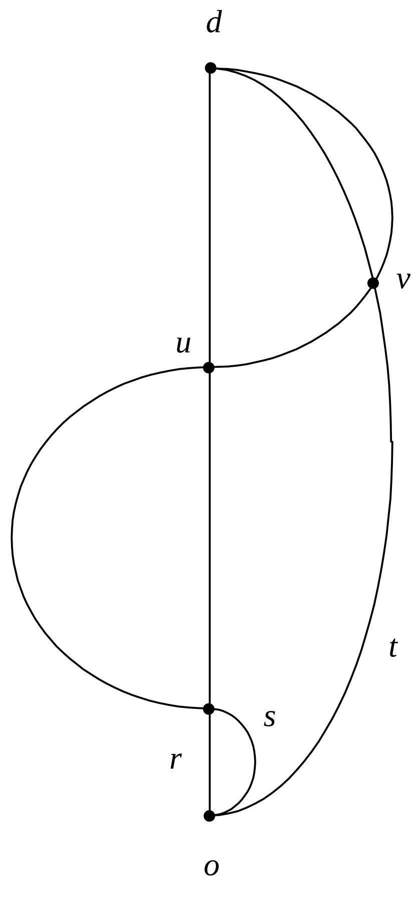


Figure 7.

Two cases are possible: Either there are two or more equivalence classes with respect to the equivalence relation in Claim 1; or this relation holds between any pair of routes in *G*. In the former case, pick up one of the equivalence classes, and consider the network *G′* that consists of all the vertices and edges that belong to at least one route in this equivalence class, as well as the network *G″* that consists of all the vertices and edges that belong to at least one route not in the class. Each vertex v, other than o and d, belongs to one, and only one, of these two networks. (Otherwise, v would belong to two routes in two different equivalence classes, which is impossible by definition of the equivalence relation.) Therefore, each edge also belongs to one, and only one, of them. This implies that G is the result of connecting G' and G'' in parallel. Clearly, since G is series-parallel, the same is true for G' and G''. Hence, G satisfies (ii).

In the rest of this proof, it will be assumed that the equivalence relation in Claim 1 holds between any pair of routes in G. It will be shown that there is some vertex, other than o and d, which is common to all routes. Note that the above assumption implies that each route in G has at least two edges. Indeed, since a route with a single edge is not equivalent to any other route, it must be the only route in G. The route's unique edge is then the only edge in G, a contradiction to the assumption that G has more than one edge.

Fix a route r and, for every route s, denote by v_s the first vertex in s, other than o, that also belongs to r. Since, by assumption, the equivalence relation holds between r and s, $v_s \neq d$. Let t be a route such that, for all routes s, v_s is either equal to v_t or precedes it in r.

CLAIM 2. The vertex v_t belongs to all routes.

Suppose the contrary, that v_t does not belong to some route s. Let u be the last vertex before v_t in r that also belongs to s. By the way t was chosen, $u \neq o$. By Lemma 1, the first vertex after u in s that also belongs to r follows v_t in r. This implies that t_{ov_t} and s_{ud} do not have a vertex in common. For if a common vertex v did exist, then the route $t_{ov} + s_{vd}$ would have the property that the vertex $v_{t_{ov}+s_{vd}}$ (i.e., the first vertex after o in which the route intersects r) follows v_t in r, which is a contradiction to the way t was chosen (see Figure 8). The fact that t_{ov_t} and s_{ud} do not have a common vertex implies that the walk $t_{ov_t} + (-r)_{v_t u} + s_{ud}$ is a route (i.e., it does not pass through any vertex more than once). However, this contradicts the assumption that G is series-parallel. This contradiction proves Claim 2.

It follows from Claim 2 that G is the result of connecting two series-parallel networks in series: the network G' that consists of v_t (as destination) and all the vertices and edges that precede it in some route in G; and the network G'' that consists of v_t (as the origin) and all the vertices and edges that follow it in some route in *G*. By Lemma 1, v_t is the only vertex common to both networks. Since *G′* and *G″* are clearly series-parallel, *G* satisfies (iii). ■

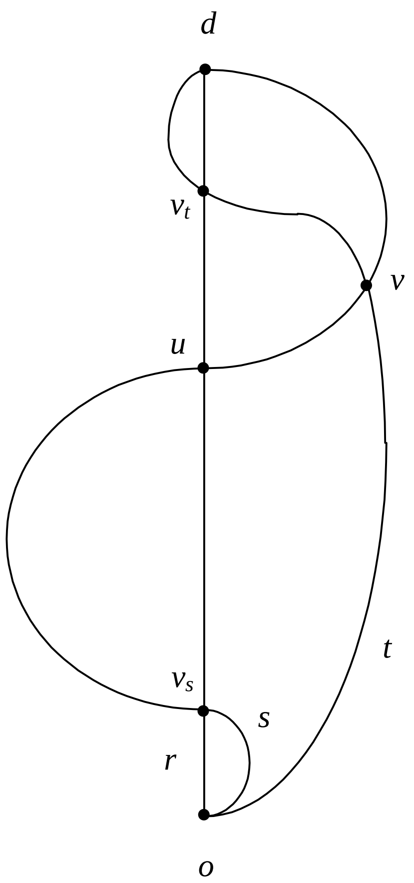


Figure 8.

Proof of Proposition 3. Since, clearly, (iii) implies (ii) and (ii) implies (i), it suffices to show that every series-parallel network *G* satisfies (iii). This will be proved by induction on the number of edges in *G*. If there is only one edge, (iii) holds trivially. Suppose, then, that *G* has more than one edge and that (iii) holds for any series-parallel network with a smaller number of edges than *G*. By Proposition 2, *G* is the result of connecting two series-parallel networks in series or in parallel. By the induction hypothesis, each of these two networks satisfies (iii). It will now be shown that this implies that *G* itself also satisfies (iii).

Condition (iii) is equivalent to the following:

> There is a one-to-one function j from the set of vertices to the integers such that, for every pair of vertices *u* and *v*, if *u* precedes *v* in some route *r* then $j(u) < j(v)$.

When two networks G' and G'' that satisfy (iii) are connected in series or in parallel, a function j' as above exists for G', and another one j'' exists for G''. It is, moreover, not difficult to see that these functions can be chosen in such a way that, for every vertex u in G' and every vertex v in G'', $j'(u) = j''(v)$ if and only if $u = v$. The unique common extension j of j' and j'' to the union of the two sets of vertices then satisfies the above condition for the network that is the result of connecting G' and G'' in series or in parallel. Hence, that network also satisfies (iii). ■

The proof of Proposition 4 uses the following lemma.

Lemma 2. For a series-parallel network G, the following three conditions are equivalent:

(i) One of the networks in Figure 2 is embedded in G.

(ii) There is a triplet of routes in G that constitutes a "bad configuration." (This term is defined in the paragraph that follows Proposition 5.)

(iii) There is a pair of routes r and s, and a vertex v common to both routes, such that $r_{ov} \neq s_{ov}$ and $r_{vd} \neq s_{vd}$.

Proof. (i) ⇒ (ii). This is clear, since a bad configuration exists in both networks in Figure 2.

(ii) ⇒ (iii). Suppose there is a bad configuration: a route t and two edges e_1 and e_2 in this route, with e_1 preceding e_2, such that e_2, but not e_1, also belongs to another route s, and e_1, but not e_2, also belongs to a third route r. Note that, since the network is series-parallel, r and s pass through e_1 and e_2, respectively, in the same directions as t. Let u be the first vertex in s that also belongs to t and, in that route, follows e_1 but precedes e_2. Let v be the last vertex in r that also belongs to t and, in that route, follows e_1 but precedes e_2 (see Figure 9). If s_{ou} and r_{vd} have a common vertex v', then (iii) holds: $r_{ov'} \neq s_{ov'}$ since e_1 belongs to $r_{ov'}$ but not to $s_{ov'}$, and $r_{v'd} \neq s_{v'd}$ since e_2 belongs to $s_{v'd}$ but not to $r_{v'd}$. Suppose, then, that s_{ou} and r_{vd} do not have a vertex in common. This implies that u is either equal to v or precedes it in r. For if u followed v in r, then the walk $s_{ou} + (-t)_{uv} + r_{vd}$ would be a route (i.e., it would not pass through any vertex more than once), which is a contradiction to the assumption that the network is series-parallel. Consider the two routes t and $s_{ou} + t_{uv} + r_{vd}$. The inequality

$t_{ov} \neq s_{ou} + t_{uv}$ holds, since e_1 belongs to the walk on the right but not to the one on the left; and $t_{vd} \neq r_{vd}$ holds, since e_2 belongs to the walk on the right but not to the one on the left. Therefore, (iii) holds, with t and $s_{ou} + t_{uv} + r_{vd}$ replacing r and s, respectively.

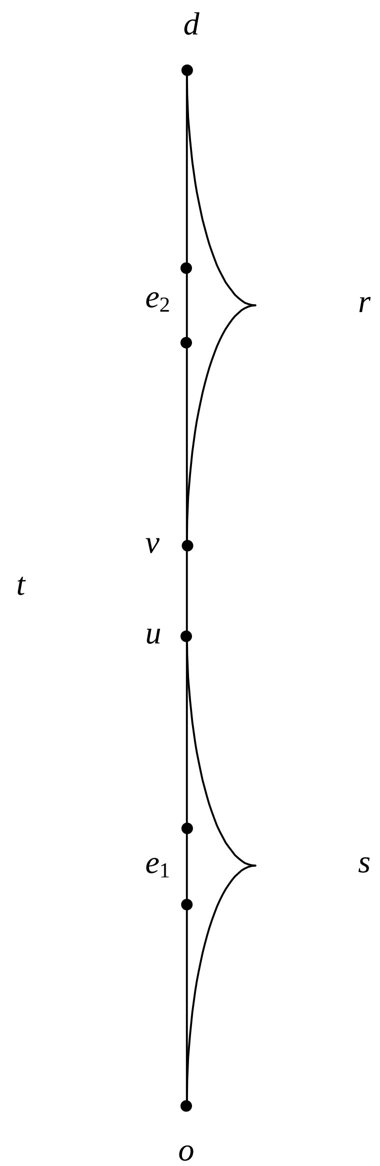


Figure 9.

(iii) ⇒ (i). Let r, s, and v be as in (iii). Let e_1 be the last edge in r that precedes v and does not belong to s, and v_1 the end vertex of e_1 which follows this edge in r. Let e_2 be the first edge in r that follows v and does not belong to s, and u_2 the end vertex of e_2 which precedes this edge in r (see Figure 10). Let u_1 be the last vertex before v_1 in s that also belongs to r, and v_2 the first vertex after u_2 in s that also belongs to r. Consider the network that consists of all the vertices and edges in r, $s_{u_1v_1}$, and $s_{u_2v_2}$.

Since only the initial and terminal vertices of each of the last two sections of *s* are in *r*, one of the networks in Figure 2 is embedded in this network. If $v_1 = u_2$, then it is the network in Figure 2(a); and if $v_1 \neq u_2$, then it is the one in Figure 2(b). ■

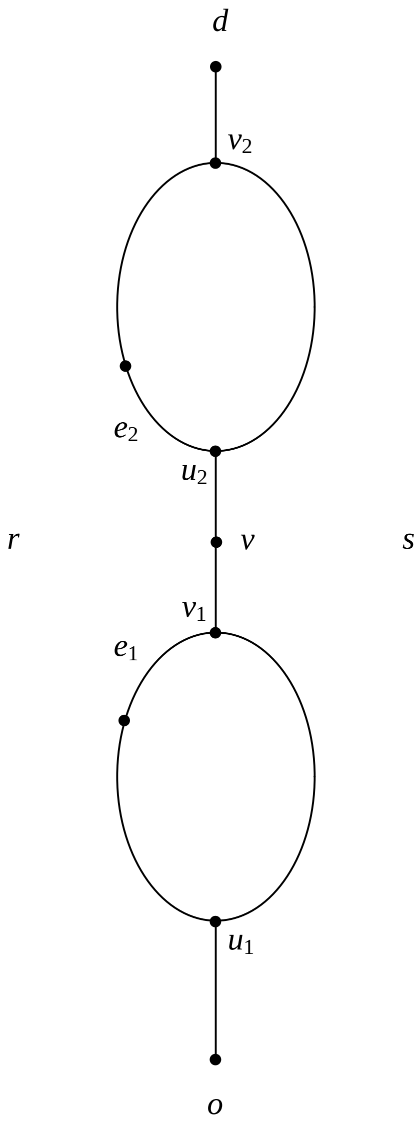


Figure 10.

Proof of Proposition 4. None of the networks in Figures 1 and 2 is a network with independent routes. Therefore, the same is true for every network *G* in which one of these networks is embedded. Conversely, if none of these networks is embedded in *G*, then, by Proposition 1, *G* is series-parallel and, by Lemma 2, a triplet of routes constituting a bad configuration does not exist. This implies that *G* is a network with independent routes. For, otherwise, there would be a route *r*, every edge of which also belongs to some other route. Clearly, no route $s \neq r$ can have <u>all</u> of *r*'s edges. Let *s* be a route with a maximal number of such edges. Let e_1 be an edge in r that is not in s, and $t \neq r$ a route containing e_1. Because of the way s was chosen, there is at least one edge e_2 common to r and s that is not in t. However, this implies that r, s, and t constitute a bad configuration, which is a contradiction. ■

Proof of Proposition 5. One direction is obvious: If a network satisfies (i), (ii), or (iii), then it is a network with independent routes. To prove the converse, assume that G is a network with independent routes. Since this is not the case for any of the networks in Figures 1 and 2, none of these networks is embedded in G. By Proposition 1, this implies that G is series-parallel. Also, by Lemma 2: (1) for every pair of routes r and s and every vertex v common to both routes, $r_{ov} = s_{ov}$ or $r_{vd} = s_{vd}$, and (2) a triplet of routes constituting a bad configuration does not exist in G. This implies the following two claims.

CLAIM 1. If two routes in G have a vertex in common, other than o and d, then their first edge is the same or their last edge is the same.

CLAIM 2. In the set of all routes in G, the relation “route r and route s have an edge in common” is an equivalence relation.

Claim 2 follows from the fact that, if route r shares an edge e_1 with route s, and s shares an edge e_2 with a third route t, then at least one of the edges e_1 and e_2 must be common to r and t, otherwise these three routes would constitute a bad configuration.

Two cases are possible: Either there are two or more equivalence classes with respect to the equivalence relation in Claim 2; or this relation holds between any pair of routes in G. In the former case, pick up one of the equivalence classes, and consider the network G' that consists of all the vertices and edges that belong to at least one route in this equivalence class, as well as the network G'' that consists of all the vertices and edges that belong to at least one route not in the class. It follows from Claim 1 and the definition of the equivalence relation that each vertex, other than o and d, belongs to one, and only one, of these two networks. Therefore, each edge also belongs to one, and only one, of them. This implies that G is the result of connecting G' and G'' in parallel. Since G is a network with independent routes, the same is clearly true for G' and G''. Hence, G satisfies (ii).

In the rest of this proof, it will be assumed that the equivalence relation holds between any pair of routes in G; i.e., any two routes have an edge in common.

CLAIM 3. All the routes in G have the same first edge, or they all have the same last edge.

To prove this claim, suppose there are two routes r and s that do not have the same last edge. Since, by assumption, they have some edge in common, it follows from Claim 1 that the first edge in r coincides with the first edge in s. Call this edge e. By a similar argument, every route t that does not have e as its first edge must have the same last edge as r. Similarly, it must have the same last edge as s. However, r and s do not have the same last edge. This contradiction proves that all routes must have e as their first edge, thus completing the proof of Claim 3.

It follows from Claim 3 that there is some edge e, with o or d as one of its end vertices, which belongs to all routes. This implies that e is the only edge incident with o or d, respectively. Therefore, either G has a single edge, or it is the result of connecting in series the network that consists of e and its two end vertices and the network that consists of all the other vertices and edges in G plus the end vertex of e which is not o or d. Clearly, the latter is a network with independent routes. Hence, G satisfies (i) or (iii). ■

Proof of Proposition 6. For each of the networks in Figures 1 and 2, none of the four conditions holds. Therefore, it follows from Proposition 4 that (i), (ii), and (iii) do not hold for any network G which is not a network with independent routes. Conversely, if G is a network with independent routes, then (i) clearly holds. And since G is series-parallel and none of the networks in Figure 2 is embedded in it, (ii) and (iii) hold by Lemma 2. ■

The following two lemmas are used in the proof of Proposition 7.

Lemma 3. Let G be a series-parallel network, and $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$ two flow vectors. If the total origin–destination flows satisfy $\hat{f}_o > 0$ and $\hat{f}_o \geq \tilde{f}_o$, then there is a route r such that, for all sections s of r of length zero or one (i.e., those containing only one vertex or only one edge), $\hat{f}_s > 0$ and $\hat{f}_s \geq \tilde{f}_s$. If $\hat{f}_o > \tilde{f}_o$, then a similar result holds with the last pair of inequalities replaced by $\hat{f}_s > \tilde{f}_s$.

Proof. The proof of the lemma proceeds by induction on the number of edges. For a network with a single edge, the result is trivial. Consider, then, a series-parallel network G with two or more edges, such that the result holds for any two flow vectors in any series-parallel network with a smaller number of edges than G (this is the induction hypothesis). By Proposition 3, G is the result of connecting two series-parallel networks, G' and G'', in series or in parallel. Consider, first, the case in which G' and G'' are connected in series, so that the destination in G', v, coincides with the origin in G''. The set $\mathcal{R}'$ of all routes in G' is then equal to $\{r_{ov} \mid r \in \mathcal{R}\}$. Every flow vector $\mathbf{f}$ in G induces a flow vector $\mathbf{f}'$ in G', which is defined by the equations $f'_{r'} = f_{r'}$ ($r' \in \mathcal{R}'$). The flow f'_s through any walk s in G' is given by an equation similar to (1), namely,

$$f'_s = \sum_{\substack{r' \in \mathcal{R}' \\ s \text{ is a section of } r'}} f'_{r'}. \tag{4}$$

It is, however, not difficult to see that, for every such s, $f'_s = f_s$. In particular, if the flow vectors $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$ satisfy the pair of inequalities $\hat{f}_o > 0$ and $\hat{f}_o \geq \tilde{f}_o$, or the inequality $\hat{f}_o > \tilde{f}_o$, then similar inequalities or inequality hold for the flow vectors $\hat{\mathbf{f}}'$ and $\tilde{\mathbf{f}}'$ they induce in G'. It then follows from the induction hypothesis that there is a route r' in G' such that, for all sections s of r' of length zero or one, $\hat{f}_s > 0$ and $\hat{f}_s \geq \tilde{f}_s$, or $\hat{f}_s > \tilde{f}_s$, respectively. By similar considerations, there is a route r'' in G'' such that similar inequalities or inequality hold for all sections s of r'' of length zero or one. Therefore, the same is true for the route $r = s' + s''$ in G. This completes the proof for the case in which G is the result of connecting the two series-parallel networks in series.

Suppose, now, that G is the result of connecting G' and G'' in parallel, so that o and d are also the origin and destination, respectively, in G' and G''. The set $\mathcal{R}$ of all routes in G is then the disjoint union of the set $\mathcal{R}'$ of all routes in G' and the set $\mathcal{R}''$ of all routes in G''. Every flow vector $\mathbf{f}$ in G induces a flow vector $\mathbf{f}'$ in G', which is defined as in the previous paragraph, and a flow vector $\mathbf{f}''$ in G'', which is defined in a similar manner. For every walk s in G', the flow f'_s through s is given by (4). If s is <u>not</u> one of the zero-length walks o and d, then $f'_s = f_s$. However, unlike in the case previously considered, f'_o and f'_d (which are both equal to $\sum_{r' \in \mathcal{R}'} f_{r'}$) are generally less than f_o and f_d, and the same is true for f''_o and f''_d. In fact, $f_o = f'_o + f''_o$. It follows from this equality that if $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$ are two flow vectors satisfying $\hat{f}_o > 0$ and $\hat{f}_o \geq \tilde{f}_o$, then $\hat{f}'_o > 0$ and $\hat{f}'_o \geq \tilde{f}'_o$, or $\hat{f}''_o > 0$ and $\hat{f}''_o \geq \tilde{f}''_o$. Then, by the induction hypothesis, there is some route r, either in G' or in G'' (and, hence, in G), such that, for all sections s of r of length zero or one, $\hat{f}_s > 0$ and $\hat{f}_s \geq \tilde{f}_s$. Similarly, if $\hat{f}_o > \tilde{f}_o$, then there is a route r in G' or in G'' such that, for all sections s of r of length zero or one, $\hat{f}_s > \tilde{f}_s$. ■

Lemma 4. Let G be a series-parallel network, $\mathbf{c}$ an additively separable cost function, and $\mathbf{f}^*$ a corresponding equilibrium. For every vertex v, there is a number $F(v)$ such that, for all routes r passing through v, if every edge in r belongs to some least cost route s, then $c_{r_{ov}}(\mathbf{f}^*) = F(v)$. In particular, for every route r that is itself a least cost route, $c_{r_{ov}}(\mathbf{f}^*) = F(v)$ for every vertex v in r. Consequently, $F(d)$ is the equilibrium cost.

Proof. Let r be a route in which every edge e belongs to some least cost route s. For every vertex v in r, the following two claims hold.

CLAIM 1. For every least cost route s passing through v, $c_{s_{ov}}(\mathbf{f}^*) \leq c_{r_{ov}}(\mathbf{f}^*)$.

Otherwise, $c_s(\mathbf{f}^*)$ would be greater than $c_{r_{ov}}(\mathbf{f}^*) + c_{s_{vd}}(\mathbf{f}^*)$, which is the cost of the route $r_{ov} + s_{vd}$, a contradiction to the assumption that s is a least cost route.

CLAIM 2. For every least cost route s passing through v, $c_{s_{ov}}(\mathbf{f}^*) \geq c_{r_{ov}}(\mathbf{f}^*)$.

This will be proved by induction on the length of r_{ov}. The induction hypothesis is that, for every vertex u that precedes v in r, Claim 2 holds with u replacing v. Suppose, by contradiction, that there is some least cost route t passing through v such that $c_{t_{ov}}(\mathbf{f}^*) < c_{r_{ov}}(\mathbf{f}^*)$. Clearly, this is possible only if $v \neq o$. Let e be the edge immediately preceding v in r, and u its other end vertex. Let s be a least cost route passing through e. By the induction hypothesis, $c_{s_{ou}}(\mathbf{f}^*) \geq c_{r_{ou}}(\mathbf{f}^*)$. Since $s_{uv} = r_{uv}$, and hence $c_{s_{uv}}(\mathbf{f}^*) = c_{r_{uv}}(\mathbf{f}^*)$, additive separability implies $c_{s_{ov}}(\mathbf{f}^*) \geq c_{r_{ov}}(\mathbf{f}^*)$, and hence $c_{s_{ov}}(\mathbf{f}^*) > c_{t_{ov}}(\mathbf{f}^*)$. However, Claim 1, when applied to t instead of r, gives $c_{s_{ov}}(\mathbf{f}^*) \leq c_{t_{ov}}(\mathbf{f}^*)$, which is a contradiction. This contradiction proves Claim 2.

For every vertex v that belongs to some least cost route s, set $F(v) = c_{s_{ov}}(\mathbf{f}^*)$. The value of $F(v)$ does not depend on the particular choice of s. Indeed, by Claims 1 and 2, if r is any route passing through v in which every edge e belongs to some least cost route, $c_{r_{ov}}(\mathbf{f}^*) = c_{s_{ov}}(\mathbf{f}^*)$. For all vertices v that do not belong to any least cost route, set $F(v) = 0$. ■

Proof of Proposition 7. Let $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$ be equilibria with respect to $\hat{\mathbf{c}}$ and $\tilde{\mathbf{c}}$, respectively, such that the total origin–destination flows satisfy $\hat{f}_o \geq \tilde{f}_o$. Since, by assumption, $\hat{c}_r(\mathbf{f}) \geq \tilde{c}_r(\mathbf{f})$ for all routes r and flow vectors $\mathbf{f}$, if $\hat{\mathbf{f}}$ is equal to $\tilde{\mathbf{f}}$ then the respective equilibrium costs $\hat{c}$ and $\tilde{c}$ satisfy $\hat{c} \geq \tilde{c}$. Suppose, then, that the two flow vectors are distinct (and, hence, $\hat{f}_o > 0$). By Lemma 3, there is some route r such that, for all sections s of r of length zero or one, $\hat{f}_s > 0$ and $\hat{f}_s \geq \tilde{f}_s$. The first inequality implies that every edge e in r belongs to some route t with $\hat{f}_t > 0$. Therefore, by the equilibrium condition (2) and Lemma 4, $\hat{c}_r(\hat{\mathbf{f}}) = \hat{c}$. The second inequality, together with the assumption that the cost functions are additively separable, implies that $\hat{c}_r(\hat{\mathbf{f}}) \geq \hat{c}_r(\tilde{\mathbf{f}})$. Since $\hat{c}_r(\tilde{\mathbf{f}}) \geq \tilde{c}_r(\tilde{\mathbf{f}})$ and, by definition, $\tilde{c}_r(\tilde{\mathbf{f}}) \geq \tilde{c}$, this proves that $\hat{c} \geq \tilde{c}$. The second assertion of the proposition immediately follows from the first. ■

Proof of Theorem 1. The implication (i) ⇒ (ii) is given by Proposition 7. To prove the reverse implication, recall that, by Proposition 1, if a network is not series-parallel, then the network in Figure 1 is embedded in it. As shown, in that particular network Braess's paradox does occur. The same is true for any network in which it is embedded. This can easily be seen by considering the following rules for assigning costs to the new edges created by the three operations defining embedding: For each of the two edges created when an existing edge is subdivided, the cost is one half that of the original edge; for an edge joining two existing vertices, the cost is infinite (or, at least, very high); and for the edge created when the origin or the destination are "extended," the cost is an arbitrary increasing function of the edge flow. ■

The following lemma is used in the proof of Theorem 2.

Lemma 5. A series-parallel network is a network with independent routes if and only if, for every pair of distinct flow vectors $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$ with equal total origin–destination flows, there is a route r such that $\hat{f}_s \geq \tilde{f}_s$ for all sections s of r, and $\hat{f}_r > \tilde{f}_r$.

Proof. For either of the a series-parallel networks in Figure 2, consider the following two flow vectors $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$. For the route r_1 passing through the edges e_1 and e_3, $\hat{f}_{r_1} = 0$

and $\tilde{f}_{r_1} = 2$; for the route r_2 passing through e_2 and e_4, $\hat{f}_{r_2} = 0$ and $\tilde{f}_{r_2} = 0$; for the route r_3 passing through e_1 and e_4, $\hat{f}_{r_3} = 1$ and $\tilde{f}_{r_3} = 0$; and for the route r_4 passing through e_2 and e_3, $\hat{f}_{r_4} = 1$ and $\tilde{f}_{r_4} = 0$. The only routes r such that $\hat{f}_r > \tilde{f}_r$ are r_3 and r_4. However, r_3 includes e_1, and r_4 includes e_3, and for these two edges, the edge flow in $\hat{\mathbf{f}}$ is less than in $\tilde{\mathbf{f}}$. In view of Propositions 1 and 4, this example proves that for any series-parallel network in which routes are not independent there is a pair of distinct flow vectors $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$ for which a route r as above does not exist.

The converse, that for every network G with independent routes, a route r as above exists for every pair of distinct flow vectors with equal total origin–destination flows, will be proved by induction on the number of edges. The induction hypothesis is that this is true for any network with independent routes and a smaller number of edges than G. Let $\hat{\mathbf{f}}$ and $\tilde{\mathbf{f}}$ be two distinct flow vectors in G with equal total origin–destination flows. Since the flow vectors themselves are distinct, G must have at least two edges. Therefore, it follows from Proposition 5 that G is the result of connecting two series-parallel networks, G' and G'', in series or in parallel. In the former case, one of these, say G'', has only one edge, e. Then, by similar arguments to those used in the proof of Lemma 3, the induction hypothesis implies that there is some route r' in G' such that $\hat{f}_s \geq \tilde{f}_s$ for all sections s of r', and $\hat{f}_{r'} \geq \tilde{f}_{r'}$. The route r in G obtained by appending e and d to r' has the same property. This is because every section s of r is (1) a section of r', or (2) the zero-length walk consisting of d alone, or (3) the result of appending e and d to some section t of r'—in which case the flows through t and s are always equal. This completes the proof for the case in which G is the result of connecting G' and G'' in series. If G is the result of connecting the two networks in parallel, then, again by similar arguments to those used in the proof of Lemma 3, it follows from the induction hypothesis that there is a route r in G' or in G'' such that $\hat{f}_s \geq \tilde{f}_s$ for all sections s of r, and $\hat{f}_r \geq \tilde{f}_r$. Since r is also a route in G, this completes the proof. ■

Proof of Theorem 2. Suppose that G is a network with independent routes. Let $\mathbf{c}$ be a cost function, $\mathbf{f^*}$ a corresponding equilibrium with positive total origin–destination flow, and c^* the equilibrium cost. If $\mathbf{f}$ is another flow vector, with the same total origin–destination flow as $\mathbf{f^*}$, then, by Lemma 5, there is some route r such that

$f_s \geq f^*_s$ for all sections s of r, and $f_r > f^*_r$. Since routes in G are independent, there is some section s of r of length one such that $f_s = f_r > f^*_r = f^*_s$. It follows that $c_r(\mathbf{f}) \geq c_r(\mathbf{f^*})$, and if $\mathbf{c}$ is increasing, then $c_r(\mathbf{f}) > c_r(\mathbf{f^*})$. Since $f_r > 0$, and, by definition, $c_r(\mathbf{f^*}) \geq c^*$, this proves that $\mathbf{f^*}$ is weakly Pareto efficient, and, moreover, is Pareto efficient if $\mathbf{c}$ is increasing.

Suppose now that routes in G are not independent. By Proposition 4, one of the networks in Figures 1 and 2 is embedded in G. As shown, for each of these three networks there is an increasing, additively separable cost function and a corresponding equilibrium that is not even weakly Pareto efficient. The same is true for every network in which one of these networks is embedded; the proof of this is based on the same arguments used in the proof of Theorem 1. ■

Proof of Theorem 3. Suppose that G is a network with independent routes. For a given assignment of cost functions to users, let s be an equilibrium, and t another strategy profile. If $\mathbf{f}(t) = \mathbf{f}(s)$, then it follows from the equilibrium condition (3) that $c^i_{t(i)}(\mathbf{f}(t)) \geq c^i_{s(i)}(\mathbf{f}(s))$ for all users i, and equality holds if $t(i) = s(i)$. Hence, condition (H) holds. Suppose, now, that $\mathbf{f}(t) \neq \mathbf{f}(s)$. By Lemma 5, there is some route r such that $f_s(t) \geq f_s(s)$ for all sections s of r, with strict inequality for r itself and, because of the independence of the routes, for some section s of r of length one. For all users i, $c_r^i(\mathbf{f}(t)) \geq c_r^i(\mathbf{f}(s)) \geq \min_{s\in \mathcal{R}} c_s^i(\mathbf{f}(s)) = c^i_{s(i)}(\mathbf{f}(s))$, and if i's cost function is increasing, then the first inequality is strict. Since $f_r(t) > 0$, the set of all users i with $t(i) = r$ has positive measure. This proves that the equilibrium s is weakly Pareto efficient. Moreover, since $f_r(t) > f_r(s)$, the set of all users i such that $t(i) = r$ but $s(i) \neq r$ also has positive measure. As shown above, for each such i, if i's cost function is increasing, then $c^i_{t(i)}(\mathbf{f}(t)) > c^i_{s(i)}(\mathbf{f}(s))$. This proves that if <u>all</u> users have increasing cost functions, then (H) holds, and hence the equilibrium s is hyper-efficient.

Suppose now that routes in G are not independent. There are increasing, additively separable cost functions for G and a corresponding Nash equilibrium that is not even weakly Pareto efficient. The proof of this is the same as the one given for Theorem 2. ■

REFERENCES

ALTMAN, E., AND KAMEDA, H. (2001). "Equilibria for multiclass routing problems in multi-agent networks," mimeo.

BECKMANN, M., MCGUIRE, C. B., AND WINSTON, C. B. (1956). *Studies in the Economics of Transportation.* New Haven, CT: Yale Univ. Press.

BELL, M. G. H., AND IIDA, Y. (1997). *Transportation Network Analysis.* Chichester, UK: Wiley.

CALVERT, B., AND KEADY, G. (1993). "Braess's paradox and power-law nonlinearities in networks," *Journal of the Australian Mathematical Society, Series B* **35**, 1–22.

COHEN, J. E., AND HOROWITZ, P. (1991). "Paradoxical behaviour of mechanical and electrical networks," *Nature* **352**, 699–701.

——, AND JEFFRIES, C. (1997). "Congestion resulting from increased capacity in single-server queuing networks," *IEEE/ACM Transactions on Networking* **5**, 305–310.

DAFERMOS, S., AND NAGURNEY, A. (1984). "On some traffic equilibrium theory paradoxes," *Transportation Research* **18B**, 101–110.

DUFFIN, R. J. (1965). "Topology of series-parallel networks," *Journal of Mathematical Analysis and Applications* **10**, 303–318.

FRANK, M. (1981). "The Braess paradox," *Mathematical Programming* **20**, 283–302.

HOLZMAN, R., AND LAW-YONE, N. (1997). "Strong equilibrium in congestion games," *Games and Economic Behavior* **21**, 85–101.

KONISHI, H. (2001). "Uniqueness of user equilibrium in transportation networks with heterogenous commuters," mimeo.

LAW-YONE, N. (1995). "Strong equilibrium in congestion games" [in Hebrew]. M.Sc. thesis, the Technion – Israel Institute of Technology.

MILCHTAICH, I. (1996). "Congestion games with individual-specific payoff functions," *Games and Economic Behavior* **13**, 111–124.

—— (2000). "Generic uniqueness of equilibrium in large crowding games," *Mathematics of Operations Research* **25**, 349–364.

—— (2001). "Social optimality and value-equivalence in a class of nonatomic congestion games," mimeo.

NAGURNEY, A. (1999). *Network Economics: A Variational Inequality Approach*, 2nd ed. Boston: Kluwer.

ORDA, A., ROM, R., AND SHIMKIN, N. (1993). "Competitive routing in multiuser communication networks," *IEEE/ACM Transactions on Networking* **5**, 510–521.

RIORDAN, J., AND SHANNON, C. E. (1942). "The number of two-terminal series-parallel networks," *Journal of Mathematics and Physics* **21**, 83–93.

SHEFFI, Y. (1985). *Urban Transportation Networks.* Englewood Cliffs, NJ: Prentice-Hall.

STEINBERG, R., AND ZANGWILL, W. I. (1983). "The prevalence of the Braess' paradox," *Transportation Science* **17**, 301–318.

VOORNEVELD, M., BORM, P., VAN MEGEN, F., TIJS, S., AND FACCHINI, G. (1999). "Congestion games and potentials reconsidered," *International Game Theory Review* **1**, 283–299.