# Braess's Paradox in Large Random Graphs*

Gregory Valiant[†] Tim Roughgarden[‡]

November 17, 2009

**Abstract**

Braess's Paradox is the counterintuitive fact that removing edges from a network with "selfish routing" can *decrease* the latency incurred by traffic in an equilibrium flow. We prove that Braess's Paradox is likely to occur in a natural random network model: with high probability, there is a traffic rate and a set of edges whose removal improves the latency of traffic in an equilibrium flow by a constant factor.

**Keywords:** Braess's Paradox; random graphs; selfish routing; traffic equilibria

*A preliminary version of this paper appeared in the Proceedings of the 7th ACM Conference on Electronic Commerce, June 2006.

[†]Computer Science Division, University of California, Berkeley. Part of this work was done while visiting Stanford University and supported in part by DARPA grant W911NF-05-1-0224. Email: gvaliant@eecs.berkeley.edu.

[‡]Department of Computer Science, Stanford University, 462 Gates Building, 353 Serra Mall, Stanford, CA 94305. This research was supported in part by ONR grant N00014-04-1-0725, DARPA grant W911NF-05-1-0224, an NSF CAREER Award, an ONR Young Investigator Award, and an Alfred P. Sloan Fellowship. Email: tim@cs.stanford.edu.

# 1 Introduction

*Braess's Paradox* is a counterintuitive phenomenon about routing traffic in a congested network; it was first discovered by Braess [5] in the first network shown in Figure 1. Assume that many small network users travel from the vertex $s$ to the vertex $t$, with each user choosing an $s$-$t$ path independently and selfishly, to minimize the delay experienced. Each edge of the network is labeled with its latency function, which describes the delay incurred by traffic on the link as a function of the amount of traffic that uses the link. We assume that the *traffic rate* — the total amount of traffic in the network — is 1. We also assume that traffic in the network reaches an "equilibrium flow", the natural outcome of "selfish routing" in which all traffic simultaneously travels along minimum-latency paths. In the (unique) equilibrium flow, all traffic uses the route $s \to v \to w \to t$ and experiences two units of latency. On the other hand, if we remove the edge $(v, w)$ to obtain the second network in Figure 1, then in the ensuing equilibrium flow half of the traffic uses each of the routes $s \to v \to t$ and $s \to w \to t$. In this equilibrium, all network users experience latency 3/2 and are thus better off than before. Thus, *removing links can improve the performance of the equilibrium flow of a selfish routing network.*

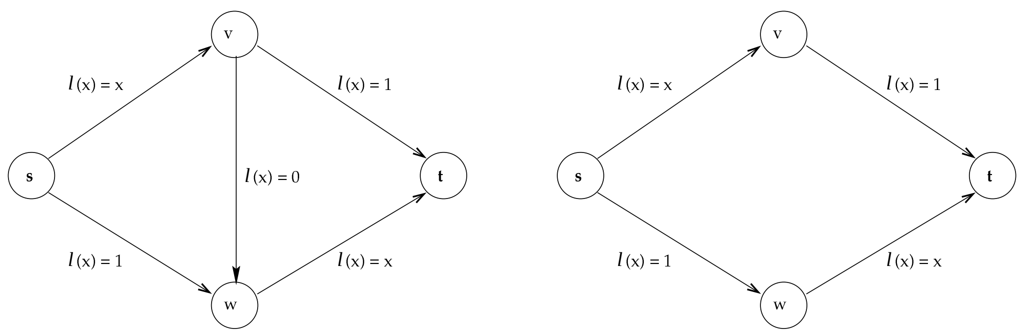



Figure 1: Braess's Paradox

Since its discovery in 1968 [5], Braess's Paradox has generated an enormous amount of research in the transportation, networking, and theoretical computer science communities (see [27] and Section 1.2 below). However, remarkably little is known about whether Braess's Paradox is a common real-world phenomenon, or a mere theoretical curiosity. Differentiating between these two possibilities is clearly an important issue. For example, it is well known that equilibrium flows arise not only in networks with "source routing"—networks where each end user is assumed to possess knowledge of the entire network and the ability to choose an end-to-end path for its traffic—but also in networks that use a distributed delay-based routing protocol to route traffic, such as the OSPF protocol with delay as the edge metric (see e.g. [3, 14]). Motivated by this fact, a recent sequence of papers in the networking literature [1, 9, 18, 19, 20] studied strategies that allocate additional capacity to a network without causing Braess's Paradox to arise—intuitively, without overprovisioning a counterproductive "cross-edge" like the edge $(v, w)$ in Figure 1. If Braess's Paradox is a rare event in selfish routing networks, then such strategies might be largely superfluous for real-world networks. If Braess's Paradox is a widespread phenomenon, then the problem of adding capacity (or new edges) to a selfish routing network must be treated with care.

In summary, the following basic but poorly understood question motivates our work:

*Is Braess's Paradox a "pathological" example or a pervasive phenomenon?*

## 1.1 Our Results

Qualitatively, our main result is the following: *in a natural random network model, Braess's Paradox occurs with high probability.* To state our results formally, define the *Braess ratio* of a network as the largest factor by which the removal of one or more edges can improve the latency of traffic in an equilibrium flow. For example, the Braess ratio of the network in Figure 1 is 4/3. For our model of random networks, we prove the following.

(R1) With high probability as $n \to \infty$, a random $n$-vertex network admits a choice of traffic rate such that the resulting Braess ratio is strictly greater than 1.

Here and throughout this paper, "with high probability" means with probability tending to 1 as $n \to \infty$. Thus Braess's Paradox is a common occurrence in large selfish routing networks, rather than an isolated anomaly.

We prove a significantly stronger result, as follows.

(R2) There is a constant $\rho > 1$ such that, with high probability as $n \to \infty$, a random $n$-vertex network admits a choice of traffic rate such that the resulting Braess ratio is at least $\rho$.

For each fixed number $n$ of network vertices, the probabilities in (R1) and (R2) are with respect to the random choice of the graph and of the edge latency functions. Our random graph model is the standard Erdös-Renyi $\mathcal{G}(n,p)$ model [10], and we work with a fairly general model of random affine latency functions (see Section 2.2 for details). The traffic rate is not random and is chosen (as a function of $n$) so that it scales appropriately with the "volume" of the network. Some such scaling of the traffic rate appears to be necessary for our results (see Section 5).

The first result (R1) already answers our motivating question and indicates that Braess's Paradox is widespread, but the second result (R2) is stronger in several respects. Most obviously, it shows that removing a set of edges can improve the latency of traffic in an equilibrium flow by a constant factor bounded away from 1 (with high probability as $n \to \infty$). We can also quantify this constant in some cases. In one model, we show that a random network typically has a Braess ratio arbitrarily close to 4/3, the largest possible in networks with affine latency functions [29].

Also, the second result (R2) requires understanding the "global" structure of a random network. Our proof of (R2) shows that, in a precise sense, a random network behaves like a modest generalization of the network in Figure 1. The first result (R1), by contrast, might plausibly be proved using only "local" arguments. For example, one could try to prove (R1) as follows: networks similar to that in Figure 1 occur sufficiently frequently as subnetworks in a random network, and perhaps under some additional (frequently met) conditions, removing the "cross-edge" of one or more such subnetworks improves the equilibrium flow. (It is not clear, however, that such a proof approach can be made to work; we do not know how to prove (R1) along such lines.) The second result (R2), which shows that a coordinated removal of a large fraction of a network's edges improves the equilibrium flow latency by a constant factor, seems unprovable by any type of local argument.

## 1.2 Related Work

Several previous works have shed some understanding on the prevalence of Braess's Paradox. On the empirical side, there has been a small amount of anecdotal evidence suggesting that Braess's Paradox has occurred in certain road networks [11, 17, 24].

On the theoretical side, a number of papers have explored the ranges of parameters under which Braess's Paradox can occur; most of these, however, confined their attention to the four-node network of Figure 1 [12, 16, 25, 26] or limited generalizations [13]. Indeed, it was only recently discovered that Braess's Paradox can be more severe in large, complex networks than in Braess's original four-node example [22, 28].

Most relevant to the present work are several papers in the transportation science literature that give analytical conditions that partially characterize whether or not a given path of edges is *improving*, in the sense that its removal will improve the equilibrium flow in the network. Steinberg and Zangwill [30] and Taguchi [31] gave the earliest (independent and incomparable) results along these lines; the former paper was subsequently generalized by Dafermos and Nagurney [7]. Such analytical characterizations reduce the problem of bounding the frequency of Braess's Paradox in a random network model to the (possibly easier) problem of bounding the likelihood that a certain inequality holds. This potential application was explicitly pointed out by Steinberg and Zangwill [30], who also noted that the form of their analytical characterization of improving routes suggested that Braess's Paradox should be common rather than rare.

The approach of [7, 30, 31] suffers from some drawbacks, however. First, the ambitious goal of analytically characterizing improving edges led to a strong extra requirement in [7, 30, 31]: the analyses in these papers explicitly assume that removing the edge(s) in question does not cause new $s$-$t$ paths to carry traffic. This assumption fails e.g. in the network of Figure 1, and it is not clear that it typically holds in large random networks. It is singled out by Steinberg and Zangwill [30, §7] as the key open issue in their analysis.

Second, even when this additional assumption holds, it is not clear that analyzing the probability that the (somewhat complex) conditions of [7, 30, 31] hold is more tractable than directly analyzing the probability that Braess's Paradox occurs. While the condition of Steinberg and Zangwill [30] suggests that this probability could be large, rigorously analyzing it in random graphs appears challenging.

Third, all of the above characterizations consider only the effects of locally modifying a network. This local approach seems incapable of proving an analogue of our second result (R2), which shows that the coordinated deletion of a large set of edges yields a constant-factor improvement in equilibrium flow latency.

In summary, we believe the present paper to be the first to explicitly define a natural probability distribution over selfish routing networks and analyze the probability that Braess's Paradox occurs, to consider non-local network modifications, and to quantify the Braess ratio in large random networks.

## 2 The Model

### 2.1 Selfish Routing Networks

We study a single-commodity flow network, described by a graph $G = (V, E)$ with a source vertex $s$ and a sink vertex $t$. We assume for convenience that all graphs are undirected, although allowing directed graphs would not affect our results in any significant way. We denote the set of simple $s$-$t$ paths by $\mathcal{P}$, and we assume that this set is nonempty. A *flow* $f$ is a nonnegative vector, indexed by $\mathcal{P}$. For a fixed flow $f$ we define $f_e = \sum_{P \in \mathcal{P}: e \in P} f_P$ as the amount of traffic using edge $e$ en route from $s$ to $t$. With respect to a finite and positive *traffic rate* $r$, a flow $f$ is said to be *feasible* if $\sum_{P \in \mathcal{P}} f_P = r$. An edge $e$ is $f$*-flow-carrying* if $f_e > 0$; a vertex other than the source or sink is $f$*-flow-carrying* if it has an $f$-flow-carrying incident edge.

We model congestion in the network by assigning each edge $e$ a nonnegative, continuous, nondecreasing *latency function* $\ell_e$ that describes the delay incurred by traffic on $e$ as a function of the edge congestion $f_e$. The latency of a path $P$ in $G$ with respect to a flow $f$ is then given by $\ell_P(f) = \sum_{e \in P} \ell_e(f_e)$. We call a triple $(G, r, \ell)$ an *instance*.

Section 1 informally discussed equilibrium flows; we now make this notion precise.

**Definition 2.1 ([32])** A flow $f$ feasible for $(G, r, \ell)$ is *at Nash equilibrium* or is a *Nash flow* if for all $P_1, P_2 \in \mathcal{P}$ with $f_{P_1} > 0$, $\ell_{P_1}(f) \leq \ell_{P_2}(f)$.

Thus all paths in use by a flow at Nash equilibrium have equal latency. Every selfish routing network admits at least one Nash flow [2]. Moreover, Nash flows are "essentially unique" in the sense that the latency incurred by traffic is the same in every Nash flow of a network [2]. We use $L(G, r, \ell)$ to denote the common latency of all traffic in a flow at Nash equilibrium for the instance $(G, r, \ell)$.

The following well-known characterization of Nash flows will be instrumental in our proofs. It follows easily from the fact that a flow at Nash equilibrium routes traffic only on minimum-latency paths.

**Proposition 2.2 ([28])** *Let $f$ be a flow feasible for $(G,\ r,\ \ell)$. For a vertex $v$, let $d(v)$ denote the length, with respect to edge lengths $\ell_e(f_e)$, of a shortest $s$-$v$ path in $G$. Then*

$$d(w) - d(v) \leq \ell_e(f_e)$$

*for all edges $e = (v, w)$, and $f$ is at Nash equilibrium if and only if equality holds whenever $f_e > 0$.*

We also use the intuitive but non-obvious fact that the latency $L(G, r, \ell)$ of traffic in a Nash flow is continuous and increasing in the traffic rate $r$.

**Proposition 2.3 ([15, 21])** *For every fixed network $G$ and strictly increasing latency functions $\ell$, the value $L(G, r, \ell)$ is continuous and strictly increasing in $r$.*

## 2.2 Models of Random Networks

To rigorously claim that Braess's Paradox is or is not likely to occur, we need to fix a model of random selfish routing networks. Such a model contains (at least) two ingredients: a probability distribution over graphs and a probability distribution over edge latency functions. While the field of random graph theory (e.g. [4]) provides many possible definitions of and analytical tools for random graphs, choices for the definition of a "random latency function" are less obvious. In this paper, we make the following two basic modeling assumptions.

(1) The underlying graph $G$ is distributed according to the standard Erdös-Renyi $\mathcal{G}(n, p)$ model [10]. Precisely, for a fixed number $n \geq 2$ of vertices, two of which are designated as a source $s$ and a sink $t$, we assume that each possible (undirected) edge is present independently with probability $p$. We also assume that $p = \Omega(n^{-1/2+\zeta})$ for some $\zeta > 0$.

(2) Latency functions are *affine*—of the form $\ell(x) = ax + b$ with $a, b \geq 0$.

We make the first assumption simply because the Erdös-Renyi model is the most popular and widely studied definition of a random graph. Our proof techniques do not crucially use detailed properties of this model, however, and we suspect that they are general enough to apply to every random graph model where a typical graph is "sufficiently dense and uniform". Whether or not our results carry over to models of sparse or non-uniform random graphs is an interesting open question.

Our motivation for assumption (2) is that affine latency functions are, informally, the simplest functions that allow Braess's Paradox to occur. More precisely, in networks with only constant latency functions or with only affine latency functions with zero constant terms, deleting edges can only increase the latency of a flow at Nash equilibrium [8]. On the other hand, allowing nonlinear latency functions only increases the worst-case severity of Braess's Paradox. For example, the network of Figure 1 has the largest-possible Braess ratio among all networks with affine latency functions [29], but larger Braess ratios are possible in networks with nonlinear latency functions [28].

Since our goal is to lower bound both the frequency and severity of Braess's Paradox in random networks, our restriction to the relatively benign class of affine latency functions is well motivated. Also, our analysis approach will be evidently robust enough to extend, with some work, to simple models of random nonlinear latency functions.

Even for affine latency functions, many models are possible. We focus most of our attention on the *independent coefficients* model (Section 3). Here, we assume that there are two fixed distributions $\mathcal{A}$ and $\mathcal{B}$, and each edge is independently given a latency function $\ell(x) = ax + b$, where $a$ and $b$ are drawn independently from $\mathcal{A}$ and $\mathcal{B}$, respectively. We prove our main result for this model — for almost every random network, for some traffic rate, removing some set of edges improves the latency of a Nash flow by a constant factor — under mild assumptions on the distributions $\mathcal{A}$ and $\mathcal{B}$.

We also consider the *$1/x$ model*, where each edge is assigned independently the latency function $\ell(x) = x$ with probability $q$ and the latency function $\ell(x) = 1$ with probability $1 - q$. This model is not a special case of the independent coefficients model, as there is now (complete) dependence between the $a$- and $b$-coefficients of the latency function of an edge. While stylized, this model serves several purposes: it shows that independence of coefficients is not essential for our earlier results; it provides a clean example of how our high-level proof approach can be adapted to different random network models; and we can obtain a precise bound on the Braess ratio of a random network in this model (as a function of the parameters $p$ and $q$). For sufficiently small values of $pq$, we prove that a random network in this model is essentially a worst-possible example of Braess's Paradox.

# 3 The Independent Coefficients Model

## 3.1 Reasonable Distributions

As discussed in Section 2.2, we assume that the underlying graph $G$ is drawn from $\mathcal{G}(n,p)$ with $p = \Omega(n^{-(1/2)+\zeta})$ for some $\zeta > 0$. We also assume that each edge latency function has the form $\ell(x) = ax + b$ where $a$ and $b$ are drawn independently from distributions $\mathcal{A}$ and $\mathcal{B}$ that satisfy some mild technical conditions. Precisely, we call $\mathcal{A}, \mathcal{B}$ *reasonable* if:

(R1) $\mathcal{A}$ has bounded support $[A_{\min}, A_{\max}]$ with $A_{\min} > 0$;

(R2) $a$-coefficients are at least somewhat dense in some closed interval $I_A$ of positive length, in the sense that for every subinterval $J \subseteq I_A$ of positive length, $\Pr[a \in J] > 0$;

(R3) $b$-coefficients are at least somewhat dense around zero in some closed interval $I_B = [0, \eta]$ with $\eta > 0$, in the sense that for every subinterval $J \subseteq I_B$ of positive length, $\Pr[b \in J] > 0$.

The assumption that $a$-coefficients are bounded away from 0 is necessary, in that otherwise a random network almost surely contains an $s$-$t$ path with essentially zero latency. Braess's Paradox will not occur in this case. The other assumptions are satisfied by most natural continuous distributions.

Our main result for the independent coefficients model is the following.

**Theorem 3.1** *Let $p = \Omega(n^{-(1/2)+\zeta})$ be an edge sampling probability with $\zeta > 0$ and $\mathcal{A}, \mathcal{B}$ reasonable distributions. There is a constant $\rho = \rho(\zeta, \mathcal{A}, \mathcal{B}) > 1$ such that, with high probability, a random network $(G, \ell)$ admits a choice of traffic rate $r$ such that the Braess ratio of the instance $(G, r, \ell)$ is at least $\rho$.*

## 3.2 Proof Approach

At the highest level, our plan is to show that a random network has a "global" structure similar to that of the four-node network of Figure 1. For a sufficiently large random network $(G, \ell)$, we choose a suitable traffic rate $r$ (scaling appropriately with the size of $G$) and consider a Nash flow $f$ for $(G, r, \ell)$. Let $d(v)$ denote the length of a shortest $s$-$v$ path with respect to the edge latencies induced by $f$. For example, in the first network of Figure 1, $d(s) = 0$, $d(v) = d(w) = 1$, and $d(t) = 2$; in the second network, $d(s) = 0$, $d(v) = 1/2$, $d(w) = 1$, and $d(t) = 3/2$. Label the $f$-flow-carrying vertices $s = v_1, \ldots, v_k = t$ so that $d(v_1) \leq \cdots \leq d(v_k)$. A key step in our analysis, which we call the "Delta Lemma", is to show that $d(v_2) \approx d(v_{k-1})$, in the sense that $d(v_{k-1}) - d(v_2) \ll d(v_2)$, with high probability. In other words, all "internal vertices" that are used by the Nash flow, $v_2, \ldots, v_{k-1}$, have relatively equal distance from the source (and the sink). Intuitively, the Delta Lemma holds because there are far more "internal" edges (edges with endpoints $v_i, v_j$, $2 \leq i, j, \leq k - 1$) than edges incident to the source and sink. We can thus regard $G$ as essentially two sets of parallel links with a small latency of $\delta = d(v_{k-1}) - d(v_2)$ associated to the center node (with respect to the flow at Nash equilibrium). We also prove an intuitive but technically non-trivial "Balance Lemma", which states that the latency of Nash flow paths is "balanced" between the two "halves" of the network, in the sense that $d(v_2), d(v_{k-1}) \approx d(t)/2$, with high probability. See Figure 2.

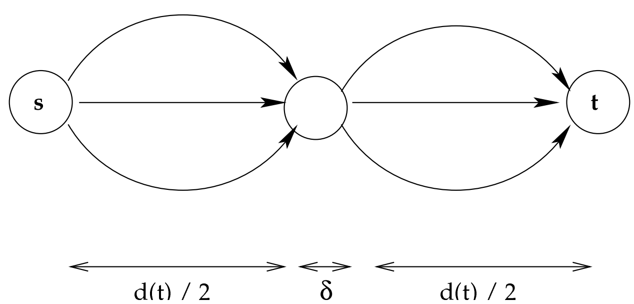
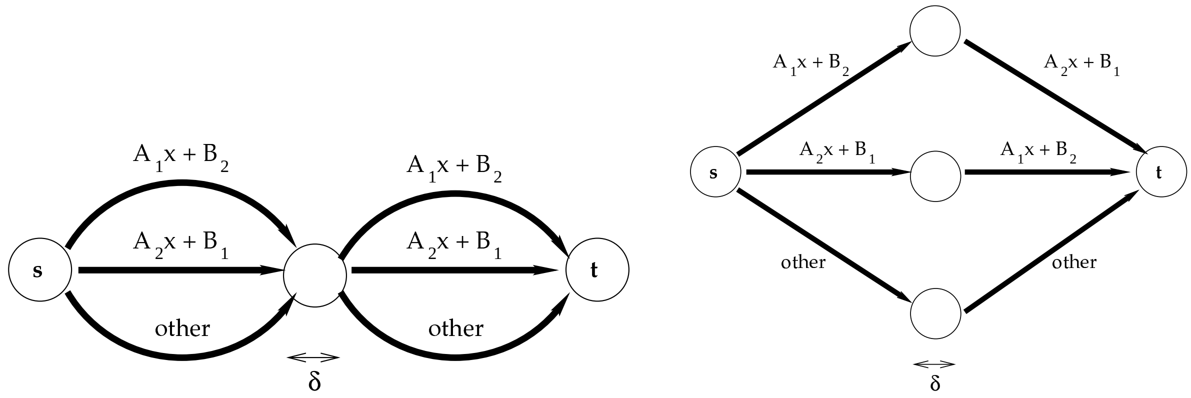



Figure 2: Delta and Balance Lemmas: A good approximation of the Nash flow latencies in a typical random network at a suitably chosen traffic rate.

Next, we partition each of the two sets of parallel links into three groups. First are the edges with a latency function with a $b$-coefficient that is roughly a parameter $B_2 > d(v_2) + \delta$ and an $a$-coefficient that is roughly a small parameter $A_1$; by Proposition 2.2 and the definition of $\delta$, these edges carry no traffic in the Nash flow of $G$. Second are the edges with a latency function with a $b$-coefficient that is roughly a constant $B_1$ that is significantly smaller than $d(v_2)$, and also an $a$-coefficient that is roughly a parameter $A_2 > A_1$. Edges in these two groups play roles analogous



(a) Original Network $G$

(b) Subnetwork $G'$

Figure 3: Partitioning the edges of a random network $G$, and the network $G'$ obtained from $G$ by deleting all "cross edges". A thick line denotes several "essentially parallel" edges.

to the edges with the latency functions $\ell(x) = 1$ and $\ell(x) = x$ in Figure 1, respectively. Third are the remaining edges. Figure 3(a) shows the network $G$ following this partitioning.

We obtain a subnetwork $G'$ by deleting edges from $G$ in order to pair up the (unused) edges with latency function roughly $A_1x + B_2$ with those with latency function roughly $A_2x + B_1$. These edge deletions are analogous to the removal of the edge $(v, w)$ in Figure 1; see Figure 3(b). Proving that this yields an improved flow at Nash equilibrium requires a careful comparison between the benefit of employing a larger number of flow paths in $G'$ and the cost of using edges with relatively large $b$-coefficients.

### 3.3 Good Networks

This section isolates a number of combinatorial properties possessed by almost all large random networks; the next three sections prove that these properties are sufficient for a Braess ratio bounded away from 1.

Fix an edge density $p = \Omega(n^{-(1/2)+\zeta})$ with $\zeta > 0$ and reasonable distributions $\mathcal{A}, \mathcal{B}$. In the notation of the defining conditions of such distributions, let $A_1 < A_2$ denote two points from the interior of the interval $I_A$ — the two points that equally trisect $I_A$, say — and let $B$ denote the midpoint of interval $I_B$. Set the parameter $\epsilon = \epsilon(\zeta, A_1, A_2)$ to be a sufficiently small positive constant; it will satisfy $(1 - \epsilon)A_2 > A_1$ (and hence $\epsilon < 1$) and its exact value will be chosen in the proof of Theorem 3.1.

Consider a random network $G$ with source $s$ and sink $t$. An edge is of 1-*type* if its $a$-coefficient is at most $A_1$ and its $b$-coefficient lies in the interval $(B, (1 + \epsilon)B)$. An edge is of $X$-*type* if its $a$-coefficient lies in the interval $((1 - \epsilon)A_2, A_2)$ and its $b$-coefficient is less than $\epsilon B$. An edge is *useless* if its $b$-coefficient is at least $(1 + \epsilon)B$. We next classify the vertices $v$ other than $s$ and $t$ into *groups*. It will be convenient to equalize the expected sizes of some of the groups. Toward this end, let $q_1$ and $q_X$ denote the probabilities that a given edge is 1- or $X$-type, respectively. Requirements (R2) and (R3) of reasonable distributions and our choices of the parameters $A_1, A_2, B, \epsilon$ ensure that both of these probabilities are lower bounded by some positive constant. If $pq_1(1 - pq_1) < pq_X(1 - pq_X)$, then the groups are:

- if $(s, v) \in G$ and is 1-type, while $(v, t)$ is either absent or not 1-type, assign $v$ to $S_1$;
- if $(s, v) \in G$ and is $X$-type, while $(v, t)$ is either absent or not $X$-type, assign $v$ to $T_1$ with probability $pq_1(1 - pq_1)/pq_X(1 - pq_X)$ and to $U$ with the remaining probability;
- if $(v, t) \in G$ and is 1-type, while $(s, v)$ is either absent or not 1-type, assign $v$ to $S_2$;
- if $(v, t) \in G$ and is $X$-type, while $(s, v)$ is either absent or not $X$-type, assign $v$ to $T_2$ with probability $pq_1(1 - pq_1)/pq_X(1 - pq_X)$ and to $U$ with the remaining probability;
- if each of $(s, v)$ and $(v, t)$ is either absent or useless, assign $v$ uniformly at random to one of the sets $Q_1, Q_2, Q_3$;
- otherwise unclassified vertices are assigned to $U$.

If $pq_1(1-pq_1) \geq pq_X(1-pq_X)$, then the assignments in the second and fourth cases are deterministic (to $T_1$ and $T_2$, respectively), while the assignments in the first and third cases are to $S_1$ and $S_2$, respectively, with probability $pq_X(1 - pq_X)/pq_1(1 - pq_1)$, and to $U$ with the remaining probability.

This (random) grouping depends on $G$ and $\ell$, on $\mathcal{A}$ and $\mathcal{B}$ (which determine $A_1, A_2, B$), and also on the choice of $\epsilon$. Observe that every vertex other than $s$ and $t$ is assigned to at least one set, and we allow a vertex to be assigned to both $S_1$ and $T_2$, or to both $S_2$ and $T_1$.

Informally, a fixed network will be called *good* if there exists an outcome of the randomized grouping experiment such that a number of random variables take on values close to their expectations, where the expectation is over both the choice of the random $n$-vertex network, and of the random grouping of the nodes of such a network. Precisely, for constants $\gamma, \tau > 0$, a network $(G, \ell)$ with $n$ vertices is $(\gamma, \tau)$-*good* if, with positive probability, randomly grouping its vertices as above causes the following random variables to take on values within a $(1 \pm (pn)^{-1/3})$ factor of their expected values for a random grouping of a random $n$-vertex network:

(P1) the sizes of all of the above eight vertex groups;

(P2) for every pair $i, j$ of nonnegative integers with $i < A_{\max}/\tau$ and $j < B/\tau$ and each vertex $v = s, t$, the number of edges $(v, w)$ with $w \in V \backslash \{s, t\}$ and with $a$- and $b$-coefficients in the intervals $[i\tau, (i + 1)\tau]$ and $[j\tau, (j + 1)\tau]$, respectively;

(P3) for every pair $i, j$ of nonnegative integers with $i < A_{\max}/\tau$ and $j < B/\tau$ and each vertex $v = s, t$, the number of edges $(v, w)$ with $w \in U$ and with $a$- and $b$-coefficients in the intervals $[i\tau, (i + 1)\tau]$ and $[j\tau, (j + 1)\tau]$, respectively;

and also the following random variables take on values within a factor 2 of their expected values for a random grouping of a random network:

(P4) for every pair $u, v$ of vertices and every value $i = 1, 2, 3$, the number of vertices $w$ in $Q_i$ such that edges $(u, w)$ and $(v, w)$ exist in $G$ and have $b$-coefficient at most $\gamma$.

Intuitively, property (P4) is useful for showing that vertices other than $s$ and $t$ are highly connected via short paths (as measured by the length parameter $\gamma$), which is central to our "Delta Lemma" (Sections 3.2 and 3.5). Properties (P2) and (P3) assert that the distributions of edge latency functions incident to the source and the sink are nearly identical (as measured by the discretization parameter $\tau$). We leverage this fact in our proof of the "Balance Lemma" (Sections 3.2 and 3.6).

For arbitrarily small constants $\gamma$ and $\tau$, a sufficiently large random network is $(\gamma, \tau)$-good with high probability.

**Lemma 3.2** *Let $p = \Omega(n^{-(1/2)+\zeta})$ be an edge sampling probability with $\zeta > 0$, $\mathcal{A}, \mathcal{B}$ reasonable distributions with parameters $A_1, A_2, B$, and $\epsilon = \epsilon(\zeta, A_1, A_2)$ a sufficiently small constant. For every pair $\gamma, \tau > 0$ of constants, a random $n$-vertex network $G$ is $(\gamma, \tau)$-good with probability approaching 1 as $n \to \infty$.*

*Proof:* We show that a random grouping of a sufficiently large random network satisfies properties (P1)–(P4) with high probability. This fact is a relatively straightforward consequence of the following Chernoff bounds (e.g. [23]): if $X$ is the sum of independent Bernoulli trials, then

$$\Pr(X \leq (1-\beta)E[X]) \leq e^{-E[X]\beta^2/2}$$

and

$$\Pr[X \geq (1+\beta)E[X]] \leq e^{-E[X]\beta^2/3}$$

for $\beta \in [0, 1]$.

In more detail, for property (P1), vertices $v$ are assigned to the various vertex groups independently, as a function only of the presence (or absence) of the edges $(s, v)$ and $(v, t)$, the latency functions of these edges, and additional coin flips specific to $v$. A vertex has probability $\Omega(p)$ of being assigned to each group, where the hidden constant depends on the distributions $\mathcal{A}, \mathcal{B}$ and the choice of $\epsilon$. This constant is bounded away from zero because of requirements (R2) and (R3) of reasonable distributions and the definitions of the parameters $A_1, A_2, B, \epsilon$. The expected size of each group is therefore $\Omega(pn) = \Omega(n^{1/2+\zeta})$, and the Chernoff bounds (with $\beta = (pn)^{-1/3}$) immediately imply that the size of each group is within a $(1 \pm (pn)^{-1/3})$ factor of its expectation with probability inverse exponential in $n$.

The proof that property (P2) holds with high probability for a sufficiently large random network is similar: since $\tau$ is constant, there are only a constant number of choices of $i$ and $j$, and the relevant random variable for each choice is either deterministically zero or has expected value $\Omega(pn)$, with the hidden constant depending on $\mathcal{A}$, $\mathcal{B}$, and $\tau$.

For property (P3), fix choices of $i, j$ and a choice of $v = s, t$. Independently for each vertex $w \in V \setminus \{s, t\}$, imagine simultaneously flipping the random coins for the potential edges $(s, w)$ and $(w, t)$. Depending on the choices of $i$ and $j$, the probability that $w$ lies in $U$ and also the edge $(v, w)$ is present with latency function coefficients in the desired intervals is either zero or $\Omega(p)$. The above Chernoff arguments now apply.

For property (P4), fix $u, v$ and a choice of $i \in \{1, 2, 3\}$. Independently for each vertex $w \in V \setminus \{s, t, u, v\}$, imagine simultaneously flipping the following independent random coins: those for the potential edges $(u, w)$ and $(v, w)$; those for the potential edges $(s, w)$ and $(w, t)$ (to determine whether or not $w \in Q_1 \cup Q_2 \cup Q_3$); and a random choice of $j \in \{1, 2, 3\}$ (to determine which, if any, of the $Q_j$'s $w$ lies in). The probability that $w \in Q_i$ is at least $\frac{1}{3}(\Pr[b > (1+\epsilon)B])^2$, which is $\Omega(1)$ by requirement (R3) and our choices of $B$ and $\epsilon$; the probability that this holds and also both $(u, w)$ and $(v, w)$ are present in $G$ with $b$-coefficients at most $\gamma$ is $\Omega(p^2)$, with the hidden constant depending on $\mathcal{B}$ and $\gamma$. Thus, the expected number of vertices $w$ that satisfy the desired properties is $\Omega(p^2 n) = \Omega(n^{2\zeta})$. The Chernoff Bounds (with $\beta = 1/2$) imply the desired concentration result with probability inverse exponential in $n$.

Since we are concerned with only polynomially many different events, each of which fails to occur with a probability that is inverse exponential in $n$, a Union Bound completes the proof. ■

### 3.4 Key Ingredients and Main Argument

Let $(G, \ell)$ be a sufficiently large $(\gamma, \tau)$-good network with $\gamma < B$. Property (P4) implies that there is at least one (indeed, many) two-hop $s$-$t$ path such that both edges have $b$-coefficient at most $\gamma$. Thus, $L(G, 0, \ell) \le 2\gamma < 2B$. By Proposition 2.3 there is a unique traffic rate $R > 0$ such that $L(G, R, \ell) = 2B$. We call this the *associated traffic rate*. We will prove that $(G, R, \ell)$ has a Braess ratio bounded away from 1. To accomplish this, obtain the *associated subnetwork* $G'$ of $G$ by retaining only the edges whose endpoints satisfy at least one of the following conditions (using the notation of Section 3.3):

- one is the source or sink, the other lies in one of $S_1, T_1, S_2, T_2, U, \{s, t\}$;
- each is in $S_1 \cup T_2 \cup Q_1$;
- each is in $S_2 \cup T_1 \cup Q_2$;
- each is in $U \cup Q_3$.

The associated network should be compared to the caricature depicted in Figure 3(b); the three "center" nodes of that figure correspond, from top to bottom, to the vertex sets $S_1 \cup T_2 \cup Q_1$, $S_2 \cup T_1 \cup Q_2$, and $U \cup Q_3$, respectively. Edges inside these sets are included in $G'$ to keep them internally highly connected.

Our proof of Theorem 3.1 hinges on two lemmas, proved in the next two sections. First, the "Delta Lemma" states that all of the "internal nodes" of a good network $G$ enjoy close proximity with respect to the Nash flow at the associated traffic rate $R$, and similarly for each of the three "vertex groups" of the associated subnetwork $G'$ for all traffic rates close to $R$.

**Lemma 3.3 (Delta Lemma)** *Fix a constant $\delta > 0$. Let $(G, \ell)$ be a sufficiently large $(\gamma, \tau)$-good network for sufficiently small constants $\gamma = \gamma(\delta, \zeta, \mathcal{A}, \mathcal{B})$ and $\tau = \tau(\delta, \zeta, \mathcal{A}, \mathcal{B})$. Let $R$ denote the associated traffic rate and $f$ a Nash flow in $(G, R, \ell)$.*

*Group vertices of $G$ so that (P1)–(P4) hold. Let $G'$ denote the associated subnetwork, $\mu = \mu(\delta, \zeta, \mathcal{A}, \mathcal{B})$ a sufficiently small constant, $R'(\mu)$ the traffic rate at which $L(G', R'(\mu), \ell) = 2B(1-\mu)$, and $f'$ a Nash flow in $(G', R'(\mu), \ell)$.*

*Define $d(v)$- and $d'(v)$-values with respect to $f$ and $f'$, respectively, as in Proposition 2.2.*

(a) *For every pair $u, v$ of $f$-flow-carrying vertices that both lie in $V \backslash \{s, t\}$, $|d(u) - d(v)| \le \delta$.*

(b) *For every pair $u, v$ of $f'$-flow-carrying vertices that both lie in $S_1 \cup T_2 \cup Q_1$, in $S_2 \cup T_1 \cup Q_2$, or in $U \cup Q_3$, $|d'(u) - d'(v)| \le \delta$.*

We also use the following corollary of the Delta Lemma for vertices of $G'$ that are not necessarily $f'$-flow carrying.

**Corollary 3.4** *With the same assumptions and notation as in Lemma 3.3: for every pair $u, v$ of vertices that both lie in $S_1 \cup T_2 \cup Q_1$, in $S_2 \cup T_1 \cup Q_2$, or in $U \cup Q_3$, $|d'(u) - d'(v)| \le 2 \max\{\gamma, \delta\}$.*

*Proof:* Fix $u$ and $v$, say in $S_1 \cup T_2 \cup Q_1$; the other two cases are the same. By property (P4) of good networks, the subgraph of $G$ induced by $S_1 \cup T_2 \cup Q_1$ contains a two-hop $u$-$v$ path $P$ in which both edges have $b$-coefficient at most $\gamma$. By definition, $P$ is also present in $G'$. Each edge $e$ of $P$ is either $f'$-flow carrying — in which case part (b) of the Delta Lemma and Proposition 2.2 imply that $\ell_e(f'_e) \le \delta$ — or not, in which case $\ell_e(f'_e) \le \gamma$. The corollary now follows immediately from Proposition 2.2. ■

The "Balance Lemma" states, roughly, that the latency of Nash flow paths in $G$, and also of Nash flow paths through $U$ in $G'$, is equally split between the two "halves" of the network.

**Lemma 3.5 (Balance Lemma)** *With the same assumptions and notation as in Lemma 3.3:*

*(a) For every $f$-flow-carrying vertex $v \in V \setminus \{s,t\}$, $d(v) \leq B + 2\delta$.*

*(b) For every vertex $v \in U$, $d'(v) \geq B(1-\mu) - 4\delta$.*

We conclude this section by showing how the Delta and Balance Lemmas, together with Lemma 3.2, imply Theorem 3.1.

*Proof of Theorem 3.1:* Define parameters $A_1, A_2, B$ for distributions $\mathcal{A}, \mathcal{B}$ as in Section 3.3 and let $\epsilon = \epsilon(\zeta, A_1, A_2)$ be a sufficiently small positive constant to be chosen later. Lemma 3.2 reduces the proof to showing that there are constants $\gamma = \gamma(\zeta, \mathcal{A}, \mathcal{B})$, $\tau = \tau(\zeta, \mathcal{A}, \mathcal{B})$, $n_0 = n_0(\zeta, \mathcal{A}, \mathcal{B})$, and $\rho_0 = \rho_0(\zeta, \mathcal{A}, \mathcal{B}) > 1$ such that the following holds: every $(\gamma, \tau)$-good network $(G, \ell)$ with at least $n_0$ vertices admits a traffic rate $r$ such that $\rho(G, r, \ell) \geq \rho_0$.

Fix $\delta = \delta(\epsilon, \zeta, \mathcal{A}, \mathcal{B})$ to be a sufficiently small constant; we choose its precise value at the end of the proof. Choose $\gamma = \gamma(\delta, \zeta, \mathcal{A}, \mathcal{B}), \tau = \tau(\delta, \zeta, \mathcal{A}, \mathcal{B})$ sufficiently small and $n_0 = n_0(\delta, \zeta, \mathcal{A}, \mathcal{B})$ sufficiently large so that the Delta and Balance Lemmas apply to every $(\gamma, \tau)$-good network with at least $n_0$ vertices, for all sufficiently small $\mu \leq \mu_0 = \mu_0(\delta, \zeta, \mathcal{A}, \mathcal{B})$. We can assume without loss that $\gamma, \tau < \delta$. Consider such a network $G$ with associated traffic rate $R$. Group the vertices as in Section 3.3 so that properties (P1)–(P4) hold, and let $G'$ denote the associated subnetwork. The proof plan is to examine the traffic rate $R'(\mu)$ at which $L(G', R'(\mu), \ell) = 2B(1-\mu)$ and show that $R'(\mu) \geq R$ for sufficiently small positive values of $\mu$. ($R'(\mu)$ is uniquely defined by Proposition 2.3 as long as $\gamma$ and $\mu$ are sufficiently small.)

Fix a sufficiently small positive value of $\mu$ and let $f$ and $f'$ denote Nash flows for $(G, R, \ell)$ and $(G', R'(\mu), \ell)$, respectively. Partition edges $(s, v)$ of $G$ into classes $E_S, E_T, E_U, E_Q$ according to whether $v$ belongs to $S_1$, $T_1$, $U$, or $Q_1 \cup Q_2 \cup Q_3$, respectively. If the edge $(s,t)$ is present in $G$, place it in $E_U$. We can write

$$R = \sum_{e \in E_S} f_e + \sum_{e \in E_T} f_e + \sum_{e \in E_U} f_e + \sum_{e \in E_Q} f_e. \tag{1}$$

Partition the $s$-$t$ paths $P$ of $G'$ into classes $\mathcal{P}_S, \mathcal{P}_T, \mathcal{P}_U$ according to the respective set $E_S, E_T, E_U$ that contains the first edge of $P$. (No edges of $E_Q$ remain in $G'$.) We can write

$$R'(\mu) = \sum_{P \in \mathcal{P}_S} f'_P + \sum_{P \in \mathcal{P}_T} f'_P + \sum_{P \in \mathcal{P}_U} f'_P. \tag{2}$$

By Proposition 2.2 and part (a) of the Balance Lemma, $\ell_e(f_e) \leq B + 2\delta$ for every $f$-flow-carrying edge $e$ incident to $s$, except possibly for the edge $(s,t)$. Since all edges of $E_S$ and $E_T$ are 1-type and $X$-type, respectively, we have

$$\sum_{e \in E_S} f_e \leq |E_S| \cdot \frac{2\delta}{A_{\min}} \tag{3}$$

and

$$\sum_{e \in E_T} f_e \leq |E_T| \cdot \frac{B + 2\delta}{(1-\epsilon)A_2}. \tag{4}$$

Also, since every edge of $E_Q$ has $b$-coefficient at least $(1+\epsilon)B$, $\sum_{e \in E_Q} f_e = 0$ provided we choose $\delta < \epsilon B/2$.

For an edge $e = (s,v)$ of $E_U \backslash \{(s,t)\}$ with $f_e > 0$, Proposition 2.2 and the Balance Lemma imply that

$$\ell_e(f_e) - \ell_e(f'_e) \leq d(v) - d'(v) \leq (B + 2\delta) - (B(1-\mu) - 4\delta) = 6\delta + \mu B.$$

Similarly, if $(s,t) \in G$, then $\ell_{st}(f_{st}) - \ell_{st}(f'_{st}) \leq 2\mu B$. Thus

$$f_e - f'_e \leq \frac{6\delta + 2\mu B}{A_{\min}}$$

for every $e \in E_U$, implying that

$$\sum_{e \in E_U} f_e - \sum_{P \in \mathcal{P}_U} f'_P = \sum_{e \in E_U} f_e - \sum_{e \in E_U} f'_e \leq |E_U| \cdot \frac{6\delta + 2\mu B}{A_{\min}}. \tag{5}$$

Next, let $F_T$ denote edges of the form $(w,t)$ with $w \in T_2$. For every (1-type) edge $(s,v) \in E_S$ and ($X$-type) edge $(w,t) \in F_T$, we can apply Proposition 2.2, Corollary 3.4 (recall $\gamma < \delta$), and the definitions of 1- and $X$-type edges to derive

$$\begin{aligned} 2B(1-\mu) &= d'(v) + (d'(w) - d'(v)) + (d'(t) - d'(w)) \\ &\leq \ell_{sv}(f'_{sv}) + 2\delta + \ell_{wt}(f'_{wt}) \\ &\leq [A_1 f'_{sv} + (1+\epsilon)B] + [A_2 f'_{wt} + \epsilon B] + 2\delta. \end{aligned}$$

Choosing $(s,v) \in E_S$ and $(w,t) \in F_T$ to minimize the right-hand side and simplifying, we obtain

$$\max\left\{\min_{e \in E_S} f'_e, \min_{e \in F_T} f'_e\right\} \geq \frac{B(1 - 2\mu - 2\epsilon) - 2\delta}{A_1 + A_2}$$

and hence

$$\sum_{P \in \mathcal{P}_S} f'_P \geq \min\left\{|E_S|, |F_T|\right\} \cdot \frac{B(1 - 2\mu - 2\epsilon) - 2\delta}{A_1 + A_2}. \tag{6}$$

By an identical argument,

$$\sum_{P \in \mathcal{P}_T} f'_P \geq \min\left\{|E_T|, |F_S|\right\} \cdot \frac{B(1 - 2\mu - 2\epsilon) - 2\delta}{A_1 + A_2}, \tag{7}$$

where $F_S$ denotes the edges $(w,t)$ with $w \in S_2$.

Combining (1)–(7) now yields

$$\begin{aligned} R'(\mu) - R \ \geq\ & (\min\{|E_S|, |F_T|\} + \min\{|E_T|, |F_S|\}) \cdot \left[\frac{B(1 - 2\mu - 2\epsilon) - 2\delta}{A_1 + A_2}\right] \\ & - |E_S| \cdot \frac{2\delta}{A_{\min}} - |E_T| \cdot \frac{B + 2\delta}{(1-\epsilon)A_2} - |E_U| \cdot \frac{6\delta + 2\mu B}{A_{\min}}. \end{aligned}$$

The randomized grouping procedure in Section 3.3 was defined so that the expected sizes of $E_S, E_T, F_S, F_T$ are equal — to $ypn$, where $y$ denotes $\min\{q_1(1 - pq_1), q_X(1 - pq_X)\}$ in the notation of Section 3.3. The expected size of $E_U$ equals $zpn$ for some other positive constant $z$ that depends only on $\zeta, \mathcal{A}, \mathcal{B}$. Writing $\nu(n) = (pn)^{-1/3}$, invoking property (P1) of good networks, and rewriting, we obtain

$$R'(\mu) - R \geq ypn\Bigg(\left[(1-\nu(n)) \cdot \frac{2B(1-2\epsilon)}{A_1 + A_2} - (1+\nu(n)) \cdot \frac{B}{(1-\epsilon)A_2}\right] \quad (8)$$

$$-\delta(1+\nu(n))\left[\frac{4}{A_1+A_2} + \frac{2}{(1-\epsilon)A_2} + \frac{2+(6z/y)}{A_{\min}}\right] \quad (9)$$

$$-2\mu B(1+\nu(n))\left[\frac{2}{A_1+A_2} + \frac{(z/y)}{A_{\min}}\right]\Bigg).$$

Rewriting the quantity in brackets in (8) as

$$B(A_1+A_2)^{-1} \cdot \left((1-\nu(n))(1-2\epsilon)\cdot 2 - \frac{1+\nu(n)}{1-\epsilon}\cdot\left(1+\frac{A_1}{A_2}\right)\right),$$

and recalling that $A_1 < A_2$, we see that for $n$ sufficiently large and $\epsilon$ sufficiently small (depending only on $\zeta$, $\mathcal{A}$, and $\mathcal{B}$), this quantity is bounded below by a positive constant $c = c(\zeta, \mathcal{A}, \mathcal{B})$. Choosing $\delta = \delta(\epsilon, \zeta, \mathcal{A}, \mathcal{B})$ sufficiently small ensures that the quantity in (9) is at least $-c/2$. There is then a choice of $\mu = \mu(\zeta, \mathcal{A}, \mathcal{B}) > 0$ that guarantees that $R'(\mu) \geq R$. Proposition 2.3 now implies that

$$\rho(G, R, \ell) \geq \frac{L(G, R, \ell)}{L(G', R'(\mu), \ell)} = \frac{1}{1-\mu},$$

completing the proof. ■

### 3.5 Proof of the Delta Lemma

*Proof of Lemma 3.3:* For part (a), sort the $f$-flow-carrying vertices $s = v_1, v_2, \ldots, v_k = t$ in nondecreasing order of $d$-values. We only need to prove that $d(v_{k-1}) - d(v_2) \leq \delta$. First, since no flow-carrying edges have zero latency, Proposition 2.2 implies that all flow entering $v_2$ does so via the edge $(s, v_2)$, and all flow departing $v_{k-1}$ uses the edge $(v_{k-1}, t)$. Since $d(t) = L(G, R, \ell) = 2B$, Proposition 2.2 implies that $\ell_e(f_e) \leq 2B$ for every flow-carrying edge $e$, and hence $f_e \leq 2B/A_{\min}$ for every edge $e$. Thus both the total flow entering $v_2$ and the total flow exiting $v_{k-1}$ is at most $2B/A_{\min}$.

Let $\kappa$ denote the number of two-hop $v_2$-$v_{k-1}$ paths whose edges both have a $b$-coefficient at most $\gamma$. Of the $\kappa$ corresponding edges exiting $v_2$, strictly more than half of them carry at most $5B/\kappa A_{\min}$ flow; similarly for the edges entering $v_{k-1}$. Applying Proposition 2.2 to two such edges $(v_2, w)$ and $(w, v_{k-1})$ with a common endpoint gives an upper bound on $d(v_{k-1}) - d(v_2)$:

$$d(v_{k-1}) - d(v_2) \leq 2 \cdot \left(A_{\max}\frac{5B}{\kappa A_{\min}} + \gamma\right). \quad (10)$$

As noted in the proof of Lemma 3.2, the expected number of two-hop paths with $b$-coefficient at most $\gamma$ between two vertices of a random network is $\Omega(n^{2\zeta})$, where the hidden constant depends on $\gamma$ and $\mathcal{B}$; this lower bound holds even if we restrict attention to the two-hop paths whose intermediate vertex lies in a prescribed set $Q_i$. Property (P4) of good networks then implies that $\kappa = \Omega(n^{2\zeta})$. Thus for $\gamma < \delta/3$ and $n$ larger than some constant $n_0 = n_0(\delta, \zeta, \mathcal{A}, \mathcal{B})$, the right-hand side of (10) is at most $\delta$, as desired.

The proof of part (b) is similar. Consider, for example, the set $S_1 \cup T_2 \cup Q_1$; the other two cases are the same. Let $v_2, v_{k-1}$ denote the $f'$-flow-carrying vertices in the set that have minimum and maximum $d'$-values, respectively. The upper bound from (a) on the amount of flow exiting $v_2$ and entering $v_{k-1}$ remains valid. The above arguments also imply that $G'$ contains $\Omega(n^{2\zeta})$ two-hop paths of the form $v_2 \to w \to v_{k-1}$, where $w \in Q_1$ and both edges have $b$-coefficient at most $\gamma$. As in part (a), this fact is enough to conclude the proof. ■

### 3.6 Proof of the Balance Lemma

*Proof of Lemma 3.5:* Consider part (a) of the Balance Lemma. We can assume that $\gamma = \gamma(\delta, \zeta, \mathcal{A}, \mathcal{B})$ is small enough and $n_0 = n_0(\delta, \zeta, \mathcal{A}, \mathcal{B})$ is large enough that the Delta Lemma applies to all $(\gamma, \tau)$-good networks with at least $n_0$ vertices, where $\tau = \tau(\delta, \zeta, \mathcal{A}, \mathcal{B})$ is a constant, smaller than $\delta$, whose exact value will be chosen later. Assume $G$ is such a network, and sort the $f$-flow-carrying vertices $s = v_1, v_2, \ldots, v_k = t$ in nondecreasing order of $d$-values. We prove that $d(v_{k-1}) \leq B + 2\delta$ provided $G$ is sufficiently large. Specifically, define $\sigma = d(v_{k-1}) - B$; we show that if $\sigma > 2\delta$, then the number of vertices $n$ of $G$ can be upper bounded by a constant that depends only on $\zeta, \mathcal{A}, \mathcal{B}, \delta, \gamma, \tau$.

The plan is to derive a contradiction by showing that an imbalance in latencies implies an imbalance between the amount of flow leaving the source and that entering the sink. This requires a discretization argument, as follows. Let $I_i$ denote the interval $[i\tau, (i+1)\tau)$ and call an edge *type* $(i, j)$ if its $a$- and $b$-coefficients lie in $I_i$ and $I_j$, respectively. Assume that $(s, t)$ is not an edge of $G$, which is without loss for the following argument. If a type $(i, j)$ edge $e$ incident to $t$ carries flow, then Proposition 2.2 and part (a) of the Delta Lemma imply that

$$\ell_e(f_e) \leq 2B - d(v_2) \leq 2B - d(v_{k-1}) + \delta = B - \sigma + \delta. \tag{11}$$

Substituting in the smallest-possible coefficients of a type $(i, j)$ edge $e$, we have

$$f_e \leq \frac{B - \sigma + \delta - j\tau}{i\tau}.$$

Analogously, if a type $(i, j)$ edge $e$ incident to $s$ carries flow, then

$$f_e \geq \frac{B + \sigma - \delta - (j+1)\tau}{(i+1)\tau}.$$

Thus the amount of flow on a flow-carrying type $(i, j)$ edge incident to $s$ exceeds that of a flow-carrying type $(i, j)$ edge incident to $t$ by at least a factor of

$$\frac{B + \sigma - \delta - (j+1)\tau}{B - \sigma + \delta - j\tau} \cdot \frac{i\tau}{(i+1)\tau} \geq \frac{B}{B - \delta} \cdot \frac{A_{\min}}{A_{\min} + \tau}, \tag{12}$$

where the inequality uses our standing assumptions that $\tau \leq \delta$ and $\sigma > 2\delta$. We can choose $\tau$ small enough that the right-hand side of (12) is at least a constant $c = c(\delta, \mathcal{A}, \mathcal{B}) > 1$.

Next, recall that (11) holds for flow-carrying edges incident to $t$; since $\sigma > 2\delta$, only edges incident to $t$ with $b$-coefficient at most $B - \delta$ are eligible to carry flow in $f$. On the other hand, we claim that *every* edge incident to $s$ with $b$-coefficient at most $B - \delta$ carries flow in $f$. To see why, let $(s, v)$ be such an edge. By property (P4) of good networks, there are edges $(v, w)$ and $(w, t)$ with $b$-coefficients at most $\gamma$. Arguing as in the proof of Corollary 3.4, $\ell_{vw}(f_{vw}) \leq \max\{\delta, \gamma\}$ and, using (11), $\ell_{wt}(f_{wt}) \leq \max\{\gamma, B - \sigma + \delta\}$. We can assume that $\gamma \leq \min\{\delta, B - \sigma + \delta\}$, so if $P$ denotes the path $s \to v \to w \to t$, we must have

$$2B = L(G, R, \ell) \leq \ell_P(f) \leq \ell_{sv}(f_{sv}) + B - \sigma + 2\delta.$$

Since $\sigma \geq 2\delta$ and $\ell_{sv}(0) < B$, we conclude that $f_{sv} > 0$.

Consider all pairs of intervals $I_i \subseteq [A_{\min}, A_{\max}]$ and $I_j \subseteq [0, B - \delta]$; these capture the coefficients of all $f$-flow-carrying edges incident to $t$, and avoid the coefficients of all non-$f$-flow-carrying edges incident to $s$. Let $\kappa_{ij}$ denote the number of (flow-carrying) edges of type $(i, j)$ incident to $s$; by property (P2) of good networks, the number of edges of this type incident to $t$ is at most a $(1 + (pn)^{-1/3})^2$ factor larger. Summing over these edge types shows that the total $f$-flow exiting the source is at least a $c/(1 + (pn)^{-1/3})^2$ factor times that entering the sink. Since these two quantities are of course equal (to $R$) and $c$ is bounded away from 1, we see that the assumption that $\sigma \geq 2\delta$ constrains the number $n$ of vertices of $G$ to be at most a constant. This completes the proof of part (a).

Part (b) follows from similar arguments. We first claim that for every $f'$-flow-carrying vertex $v \in U$, $d'(v) \geq B(1-\mu) - 2\delta$ (provided $G$ is sufficiently large). This follows from the proof of part (a), applied to $f'$ restricted to the subgraph of $G'$ induced by $U \cup Q_3 \cup \{s, t\}$, with minor modifications: the roles of $s$ and $t$ are exchanged, the factors of $B$ are replaced by factors of $B(1 - \mu)$, part (b) of the Delta Lemma takes the place of part (a), and property (P3) of good networks substitutes for (P2). Since $\gamma < \delta$, part (b) of the lemma follows immediately from this claim and Corollary 3.4. ■

## 4 The $1/x$ Model

This section studies the $1/x$ model introduced in Section 2.2. By a random network from $\mathcal{G}(n, p, q)$, we mean a random graph $G$ from the distribution $\mathcal{G}(n, p)$ for which each edge of $G$ is independently assigned the latency function $\ell(x) = x$ with probability $q \in (0, 1)$ and the latency function $\ell(x) = 1$ with probability $1 - q$. We call these two types of edges *$x$-edges* and *1-edges*, respectively. For simplicity, we explicitly disallow a direct $(s, t)$ edge in this section; if such an edge is present and has constant latency, Braess's Paradox will not occur. To streamline the exposition, we also assume that $p$ and $q$ are constants; it will be clear that the same proofs are valid provided $p \cdot \min\{q, 1 - q\}$ does not decrease too rapidly as a function of $n$. Our main result is then the following.

**Theorem 4.1** *Let $p, q, \epsilon \in (0, 1)$ be constants. With high probability, a sufficiently large random network $(G, \ell)$ from $\mathcal{G}(n, p, q)$ admits a choice of traffic rate $r$ such that the Braess ratio of the instance $(G, r, \ell)$ is at least*

$$\frac{4 - 3pq}{3 - 2pq} - \epsilon.$$

Observe that for small values of $pq$, the Braess ratio in Theorem 4.1 is nearly 4/3, the largest-possible Braess ratio in networks with affine latency functions [28, 29].

The proof of Theorem 4.1 follows the general approach outlined in Section 3.2, although the specifics differ. We start by identifying convenient combinatorial properties possessed by almost all large random networks from $\mathcal{G}(n, p, q)$. First, partition the vertices $v \neq s, t$ of a network $(G, \ell)$ into the following groups:

- if $(s, v)$ is a 1-edge and $(v, t)$ is an $x$-edge, put $v$ in $S_1$;
- if $(s, v)$ is not an edge and $(v, t)$ is an $x$-edge, put $v$ in $S_2$;
- if $(s, v)$ is an $x$-edge and $(v, t)$ is a 1-edge, put $v$ in $T_1$;
- if $(s, v)$ is an $x$-edge and $(v, t)$ is not an edge, put $v$ in $T_2$;

- if $(s, v)$ and $(v, t)$ are both $x$-edges, put $v$ in $U$;
- if $(s, v)$ and $(v, t)$ are both 1-edges, put $v$ in $Q_1$;
- if at least one of $(s, v), (v, t)$ is not an edge and neither is an $x$-edge, put $v$ in $Q_2$.

We call a network $(G, \ell)$ *good* if the following random variables take on values within a $(1 \pm n^{-1/3})$ factor of their expected values for a random $n$-vertex network:

(Q1) the sizes of all of the above seven vertex groups;

(Q2) for every pair $u, v$ of vertices, the number of vertices $w$ in $V \backslash \{s, t\}$ such that edges $(u, w)$ and $(v, w)$ exist in $G$ and are $x$-edges;

(Q3) for each vertex $v$, the number of $x$-edges $(v, w)$ with $w \in S_1$;

(Q4) for each vertex $v$, the number of $x$-edges $(v, w)$ with $w \in T_1$.

For arbitrarily small constants $p$ and $q$, a sufficiently large random network from $\mathcal{G}(n, p, q)$ is good with high probability. We omit the proof, which is similar to (and simpler than) that of Lemma 3.2.

**Lemma 4.2** *For every pair $p, q > 0$ of constants, a random $n$-vertex network $G$ of $\mathcal{G}(n, p, q)$ is good with probability approaching 1 as $n \to \infty$.*

We can now prove Theorem 4.1.

*Proof of Theorem 4.1:* Lemma 4.2 reduces the proof to showing that there is a constant $n_0 = n_0(p, q, \epsilon)$ such that, for every good network $(G, \ell)$ with at least $n_0$ vertices, there is a traffic rate $r$ such that $\rho(G, r, \ell) \geq (4 - 3pq)/(3 - 2pq) - \epsilon$.

Fix a good network $(G, \ell)$ with $n$ vertices. We set $r = pqn$, intuitively to "saturate" the $x$-edges leaving the source vertex. We first claim that

$$L(G, r, \ell) \geq 2 - \delta_1, \tag{13}$$

where $\delta_1 \to 0$ as $n \to \infty$. Let $f$ denote the Nash flow of $(G, r, \ell)$, let $v_2$ and $v_{k-1}$ denote the $f$-flow-carrying vertices other than $s, t$ with minimum and maximum $d$-values (in the sense of Proposition 2.2), respectively, and write $\nu(n) = n^{-1/3}$. Write

$$d(t) = (d(v_{k-1}) - d(s)) + (d(t) - d(v_2)) - (d(v_{k-1}) - d(v_2)). \tag{14}$$

For each of $v = s, t$, property (Q1) of good networks implies that there are at most $pqn(1 + \nu(n))$ $x$-edges incident to $v$. Since $r = pqn$, there is an $f$-flow-carrying edge $(v, w)$ with $\ell_e(f_e) \geq (1 + \nu(n))^{-1}$. This implies that the first two terms on the right-hand side of (14) are both at least $(1 + \nu(n))^{-1}$. To complete the claim, we can argue as in the proof of the Delta Lemma (Lemma 3.3) that the third term is $o(1)$. In detail, assume that $d(t) < 1 + (1 + \nu(n))^{-1}$; otherwise the claim is complete. In this case, $d(v_2)$ and $d(t) - d(v_{k-1})$ are both less than 1, implying that $(s, v_2)$ and $(v_{k-1}, t)$ are both $x$-edges. Moreover, at most one unit of flow enters $v_2$ (necessarily from $s$) and exits $v_{k-1}$ (necessarily to $t$). Property (Q2) of good networks guarantees that the number $\kappa$ of two-hop $v_2$-$v_{k-1}$ paths that comprise only $x$-edges satisfies $\kappa \geq (1 - \nu(n))p^2q^2n$. On some such path, both edges carry at most $3/\kappa$ units of flow, and Proposition 2.2 then implies that $d(v_{k-1}) - d(v_2) \leq 6/\kappa = o(1)$, as desired.

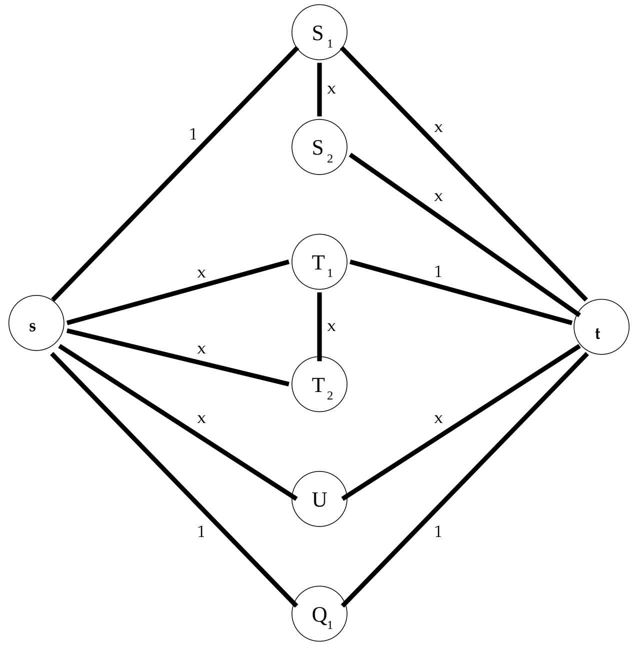



Figure 4: Proof of Theorem 4.1: structure of the subnetwork $G'$. A thick line denotes several "essentially parallel" edges.

Define a subnetwork $G'$ of $G$ by retaining only the edges incident to $s$ and $t$, and also the $x$-edges with one endpoint each in $S_1$ and $S_2$, or in $T_1$ and $T_2$ (Figure 4). We require an analog of part (b) of the Delta Lemma, as follows. Let $f'$ denote the Nash flow of $(G', r, \ell)$ and $d'$ the corresponding distance labels. We claim that if $u, v$ both lie in $S_1 \cup S_2$ or in $T_1 \cup T_2$, then $|d'(u) - d'(v)| = o(1)$. To prove this, consider the sets $T_1$ and $T_2$; the case of $S_1$ and $S_2$ is symmetric. In $G'$, all vertices of $T_1 \cup T_2$ are directly connected to $s$ and have no incident edges leading outside $T_1 \cup T_2 \cup \{s, t\}$; and each vertex of $T_1$ is directly connected to $t$ via a 1-edge. We can assume that $L(G', r, \ell) > 1$, in which case Proposition 2.2 and the structure of $G'$ imply that every ($x$-)edge between $s$ and $T_1$ carries $f'$-flow. Since no edge has zero latency and all edges between $T_1$ and $t$ are 1-edges, Proposition 2.2 implies that $f'$-flow at a vertex $v \in T_1$ travels directly to $t$. Thus, all edges between $T_1$ and $t$ are $f'$-flow-carrying. Proposition 2.2 then implies that $d'(v) = d'(t) - 1$ for every $v \in T_1$. This is also the maximum $d'$-value among vertices of $T_1 \cup T_2$. Now consider the vertex $v$ of $T_1 \cup T_2$ with minimum $d'$-value; necessarily, $v \in T_2$. The only flow entering $v$ arrives from $s$. By property (Q1) of good networks, $G'$ contains a two-hop $s$-$t$ path of 1-edges; thus $d'(t) \leq 2$. Since $(s, v)$ is an $x$-edge, $f'_{sv} \leq 2$. By property (Q4) of good networks and the definition of $G'$, $v$ is adjacent, via an $x$-edge in $G'$, to $\kappa \geq (1 - \nu(n)) \cdot p^3 q^2 (1 - q) n$ vertices of $T_1$. Thus $f'_{vw} \leq 2/\kappa$ for some edge $(v, w)$ with $w \in T_2$, and Proposition 2.2 implies that $d'(v) \geq d'(w) - 2/\kappa = (d'(t) - 1) - o(1)$, as desired.

To derive an upper bound on $L(G', r, \ell)$, let $r_S$, $r_T$, and $r_U$ denote the amount of $f'$-flow that departs the source for a vertex in $S_1 \cup S_2$, $T_1 \cup T_2$, and $U$, respectively. Since each vertex of $U$ participates in a unique $s$-$t$ path, which comprises two $x$-edges,

$$r_U = |U| \cdot \frac{L(G', r, \ell)}{2}. \tag{15}$$

Our claim that $d'(v) \geq L(G', r, \ell) - 1 - o(1)$ for every $v \in T_1 \cup T_2$ implies that

$$r_T \geq (|T_1| + |T_2|) \cdot \big(L(G', r, \ell) - 1 - \delta_2\big)\,, \tag{16}$$

where $\delta_2 = o(1)$. Symmetrically,

$$r_S \geq (|S_1| + |S_2|) \cdot \big(L(G', r, \ell) - 1 - \delta_2\big)\,. \tag{17}$$

Property (Q1) of good networks implies that $|U| \geq (1-\nu(n)) \cdot p^2 q^2 n$ and that $|S_1|+|S_2|, |T_1|+|T_2| \geq (1 - \nu(n)) \cdot pq(1 - pq)n$. Also, $r_S + r_T + r_U \leq r = pqn$. Combining these facts with (15)–(17) and rewriting, we obtain

$$L(G', r, \ell) \leq \frac{6 - 4pq + 4\delta_2(1 - pq)}{(4 - 3pq)(1 - \nu(n))}. \tag{18}$$

Since $\delta_1, \delta_2, \nu(n) = o(1)$, taking the ratio of (13) and (18) completes the proof. ■

## 5 Discussion

In our key results (Theorems 3.1 and 4.1), we assume that the graph and edge latency functions are random while the traffic rate is adversarially chosen. One could also consider random traffic rates, but there is healthy evidence that Braess's Paradox is unlikely to occur across a wide range of traffic rates (see [14, 25]). Results such as ours seem to require a carefully chosen traffic rate, although our proofs do permit a limited amount of flexibility in this choice.

We suspect that our main results continue to hold in the $\mathcal{G}(n,p)$ model even when $p = O(1/\sqrt{n})$, as well as in other random network models. A particularly interesting direction for additional research would be to consider random graph models such as power-law graphs that closely resemble naturally arising networks; in these asymmetric networks, Braess's paradox might only be present or severe when the source and sink have relatively low degree. We leave exploration of such models to future work.

## References

[1] E. Altman, R. El Azouzi, and O. Pourtallier. Avoiding paradoxes in multi-agent competitive routing. *Computer Networks*, 43(2):133–146, 2003.

[2] M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation.* Yale University Press, 1956.

[3] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods.* Prentice-Hall, 1989. Second Edition, Athena Scientific, 1997.

[4] B. Bollobás. *Random Graphs.* Academic Press, 1985.

[5] D. Braess. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, 12:258–268, 1968. English translation in [6].

[6] D. Braess. On a paradox of traffic planning. *Transportation Science*, 39(4):446–450, 2005.

[7] S. C. Dafermos and A. Nagurney. On some traffic equilibrium theory paradoxes. *Transportation Research, Series B*, 18(2):101–110, 1984.

[8] S. C. Dafermos and F. T. Sparrow. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards, Series B*, 73(2):91–118, 1969.

[9] R. El Azouzi, E. Altman, and O. Pourtallier. Properties of equilibria in competitive routing with several user types. In *Proceedings of the 41st IEEE Conference on Decision and Control*, volume 4, pages 3646–3651, 2002.

[10] P. Erdös and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:17–61, 1960.

[11] C. Fisk and S. Pallottino. Empirical evidence for equilibrium paradoxes with implications for optimal planning strategies. *Transportation Research, Part A*, 15(3):245–248, 1981.

[12] M. Frank. The Braess Paradox. *Mathematical Programming*, 20(3):283–302, 1981.

[13] M. Frank. Cost-deceptive links on ladder networks. *Methods of Operations Research*, 45:75–86, 1983.

[14] E. J. Friedman. Genericity and congestion control in selfish routing. In *Proceedings of the 43rd Annual IEEE Conference on Decision and Control (CDC)*, pages 4667–4672, 2004.

[15] M. A. Hall. Properties of the equilibrium state in transportation networks. *Transportation Science*, 12(3):208–216, 1978.

[16] H. Kameda. How harmful the paradox can be in the Braess/Cohen-Kelly-Jeffries networks. In *Proceedings of the 21st INFOCOM Conference*, volume 1, pages 437–445, 2002.

[17] G. Kolata. What if they closed 42nd Street and nobody noticed? *New York Times*, page 38, December 25, 1990.

[18] Y. A. Korilis, A. A. Lazar, and A. Orda. Capacity allocation under noncooperative routing. *IEEE Transactions on Automatic Control*, 42(3):309–325, 1997.

[19] Y. A. Korilis, A. A. Lazar, and A. Orda. Avoiding the Braess paradox in noncooperative networks. *Journal of Applied Probability*, 36(1):211–222, 1999.

[20] L. Libman and A. Orda. The designer's perspective to atomic noncooperative networks. *IEEE/ACM Transactions on Networking*, 7(6):875–884, 1999.

[21] H. Lin, T. Roughgarden, and É. Tardos. A stronger bound on Braess's Paradox. In *Proceedings of the 15th Annual Symposium on Discrete Algorithms*, pages 333–334, 2004.

[22] H. Lin, T. Roughgarden, É. Tardos, and A. Walkover. Braess's Paradox, Fibonacci numbers, and exponential inapproximability. In *Proceedings of the 32nd Annual International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 3580 of *Lecture Notes in Computer Science*, pages 497–512, 2005.

[23] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[24] J. D. Murchland. Braess's paradox of traffic flow. *Transportation Research*, 4(4):391–394, 1970.

[25] E. I. Pas and S. L. Principio. Braess' paradox: Some new insights. *Transportation Research, Series B*, 31(3):265–276, 1997.

[26] C. M. Penchina. Braess paradox: Maximum penalty in a minimal critical network. *Transportation Research, Series A*, 31(5):379–388, 1997.

[27] T. Roughgarden. *Selfish Routing and the Price of Anarchy.* MIT Press, 2005.

[28] T. Roughgarden. On the severity of Braess's Paradox: Designing networks for selfish users is hard. *Journal of Computer and System Sciences*, 72(5):922–953, 2006.

[29] T. Roughgarden and É. Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, 2002.

[30] R. Steinberg and W. I. Zangwill. The prevalence of Braess' Paradox. *Transportation Science*, 17(3):301–318, 1983.

[31] A. Taguchi. Braess' paradox in a two-terminal transportation network. *Journal of the Operations Research Society of Japan*, 25(4):376–388, 1982.

[32] J. G. Wardrop. Some theoretical aspects of road traffic research. In *Proceedings of the Institute of Civil Engineers, Pt. II*, volume 1, pages 325–378, 1952.