Journal of Automated Reasoning manuscript No.
(will be inserted by the editor)

A Mechanized Proof of the Max-Flow Min-Cut Theorem for Countable Networks with Applications to Probability Theory

Andreas Lochbihler

Received: date / Accepted: date

Abstract Aharoni et al. [3] proved the max-flow min-cut theorem for countable networks, namely that in every countable network with finite edge capacities, there exists a flow and a cut such that the flow saturates all outgoing edges of the cut and is zero on all incoming edges. In this paper, we formalize their proof in Isabelle/HOL and thereby identify and fix several problems with their proof. We also provide a simpler proof for networks where the total outgoing capacity of all vertices other than the source and the sink is finite. This proof is based on the max-flow min-cut theorem for finite networks. As a use case, we formalize a characterization theorem for relation lifting on discrete probability distributions and two of its applications.

Keywords flow network, optimization, infinite graph, Isabelle/HOL, probability

1 Introduction

The max-flow min-cut theorem for finite networks [16] has wide-spread applications: network analysis, optimization, scheduling, etc. Aharoni et al. [3] have generalized this theorem to countable networks, i.e., graphs with countably many vertices and edges, as follows:

Theorem 1 *Let $\Delta = (V, E, s, t, c)$ be a directed graph with countably many edges $E \subseteq V \times V$, vertices s and t, and a capacity function $c :: E \to \mathbb{R}_{\geq 0}$. There exists a flow f and an s-t-cut C such that f saturates all outgoing edges e of C, i.e. $f(e) = c(e)$, and is 0 on all incoming edges.*

The countable max-flow min-cut theorem is used, e.g., in probability [33] and programming language theory [26], privacy [9], and for random walks [32]. Here, we formalize this theorem in Isabelle, along with further applications.

Traditionally, the max-flow min-cut theorem is stated in terms of equality of values: The value of the maximum flow is equal to the value of the minimum cut. Here, a flow $f :: E \Rightarrow \mathbb{R}_{\geq 0}$ assigns values to the edges of Δ such that the

Digital Assert (Switzerland) GmbH, Zurich, Switzerland, E-mail: mail@andreas-lochbihler.de

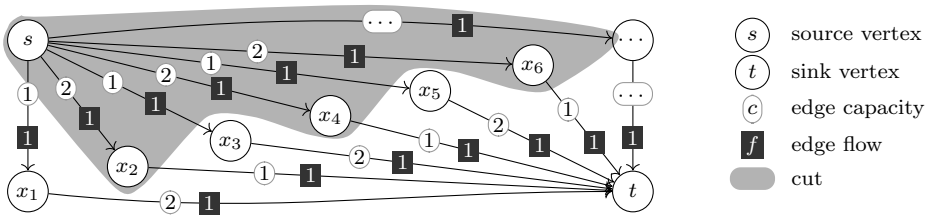

Fig. 1 A countable network with a flow and a cut of infinite value.

incoming and outgoing amounts in every vertex are the same, except for the source s and the sink t. The value $|f|$ is the amount that leaves the source s, i.e., $|f| = \sum_{x \in \mathrm{OUT}(s)} f(s,x)$ where $\mathrm{OUT}(x) = \{y \mid (x,y) \in E\}$. Dually, an s-t-cut partitions the vertices into two sets $(C, V - C)$ such that C contains the source s but not the sink t. Its value $|C|$ is the total capacity of the edges that leave C: $|C| = \sum_{e \in \mathrm{OUT}(C)} c(e)$ where $\mathrm{OUT}(C) = \{(x,y) \in E \mid x \in C \wedge y \notin C\}$.

For finite networks, the equality-of-values condition $|f| = |C|$ is equivalent to the flow f saturating the cut C. In infinite networks, the saturation condition is preferable. For example, Fig. 1 shows a network with source s and sink t and countably many vertices x_i. The edge capacities are given as white rounded rectangles on the edges. The black rectangles denote a flow f and the vertices in the grey area form a cut C. The flow f saturates the outgoing edges of C and we have $|f| = \infty = |C|$. However, there is another flow g given by $g(e) = {}^1\!/_2 f(e)$ that sends only half the amount of f. Still, $|g| = \infty = |C|$. So the equality-of-values condition does not distinguish between f and g. Yet, we should consider only f a maximum flow, not g, as one can obviously increase g on some edges. The cut-saturation condition achieves this as it compares the finite capacities of individual edges with the flow through them.

This subtlety highlights the main challenge in proving the max-flow min-cut theorem for countable networks: avoiding infinite summations. Aharoni et al.'s proof performs an elaborate dance around this problem, transforming the network several times on the way. Our formalization follows these steps through all the transformations (Sect. 3) until the problem is reduced to finding a matching in an infinite bipartite graph. The original proof then jumps back to arbitrary networks. Our proof forks into two proofs: The first takes a shortcut to a significantly simpler argument based on the max-flow min-cut theorem for finite networks (Sect. 4.1). This shortcut works only for networks where the sum of the capacities of the outgoing edges of any vertex other than the source and the sink is finite. This condition is met in some applications [9,26]. The second proof follows the original (Sect. 4.2).

As a use case, we formalize a characterization theorem for relation lifting over discrete probability distributions following Sack and Zhang [33] (Sect. 7). We apply the characterization to prove two properties: First, the functor law for discrete probabilities over the category of relations, which is key in justifying datatype recursion through discrete probability distributions [11]. Second, parametricity of the fixpoint operator on discrete subprobabilities, which is crucial in showing that a probabilistic programming language is parametric [26].

Our main contributions are as follows:

- We formalize Aharoni et al.'s strong version of the max-flow min-cut theorem for countable networks in Isabelle/HOL. The formalization has clarified the definitions and theorems and has revealed several problems in the original proofs (Sect. 6), which we have fixed. In particular, the reduction to bipartite graphs did not work as expected and required more general theorems.
- We give a considerably simpler proof for the case when every inner vertex of a network has only finite total outgoing capacity. This local boundedness assumption allows us to reuse Lammich and Sefidgar's formalization of the max-flow min-cut theorem for finite networks [23] by applying a majorised convergence argument.
- We formalize for the first time the characterization theorem for relation lifting over discrete probability distributions. Sack and Zhang's [33] proof needs the general max-flow min-cut theorem. We tweak their proof a bit so that the bounded version of the max-flow min-cut theorem suffices. This small twist arguably simplifies the original proof by removing three case distinctions. Moreover, we give short proofs of two properties about discrete probabilities using the characterization. Other formalizations already build on these properties [17, 26, 27].

Neither of the two max-flow min-cut proofs requires a large background theory; basic notions like infinite summations, monotone and majorised convergence, and fixpoints of increasing functions suffice. The formalization therefore does not rely on specific Isabelle/HOL features and could have been done similarly in other systems like HOL4 and Coq.

This paper first presents the corrected proof using conventional mathematical notation (Sects. 2–4). Informal proofs for the theorem and lemmas can be found in the accompanying report [29]. We discuss the formalization aspects in Sect. 5 and the problems with the original proof in Sect. 6. The lifting characterization and its applications are presented in Sect. 7.

The formalization of the max-flow min-cut theorem started in 2015 and a first version was published in the Archive of Formal Proofs in 2016. Unless noted otherwise, this paper describes the cleaned-up version for Isabelle2021 [25], which also includes the simpler proof for the bounded case. An earlier version of this paper has been published in [28]. This paper additionally includes the applications (Sect. 7) and discusses more of the problems we have found in Aharoni et al.'s proof (Sect. 6).

2 Graphs, Networks, and Webs

In this section, we introduce the relevant notions for graphs, networks, and webs. The terminology and notation follows [3] to ease the comparison and make the presentation accessible to mathematicians. Formalization considerations will be discussed in Sect. 5.

Definition 1 (Graph) A *(directed) graph* $G = (V, E)$ consists of a set of vertices V and a set of directed edges $E \subseteq V \times V$. A graph is countable iff its set of edges is countable. The neighbours of a vertex $x \in V$ are given by $\mathrm{OUT}_G(x) = \{\, y \mid (x, y) \in E \,\}$ and $\mathrm{IN}_G(x) = \{\, y \mid (y, x) \in E \,\}$. If the graph G is obvious from the context, we drop the subscript G.

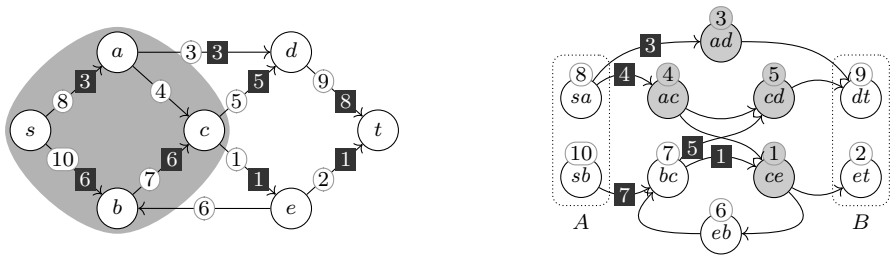

Fig. 2 Example of a network (left) and a flow (values of 0 are omitted) with an orthogonal cut, and the corresponding web (right) with a maximal wave (black rectangles) and its set of terminal vertices (grey circles). Capacities and weights are shown as labels in rounded rectangles.

Given a function $f :: E \to \mathbb{R}_{\geq 0}$, the *in-degree* $d_f^- :: V \to \mathbb{R}^\infty_{\geq 0}$ of f given by $d_f^-(x) = \sum_{y \in \mathrm{IN}(x)} f(y, x)$ assigns to each vertex $x \in V$ the sum of f over all incoming edges to x. Analogously, $d_f^+(x) = \sum_{y \in \mathrm{OUT}(x)} f(x, y)$ denotes f's *out-degree* of $x \in V$. If $d_f^+(x) = 0$, then x is a *sink* for f. The set $\mathrm{SINK}(f)$ denotes the set of sinks for f.

Definition 2 (Network) A *network* $\Delta = (V, E, s, t, c)$ is a graph (V, E) with two dedicated vertices, the source s and the sink t, and a capacity function $c :: E \to \mathbb{R}_{\geq 0}$. A network is countable iff the graph is countable.

Definition 3 (Flow) For a network $\Delta = (V, E, s, t, c)$, a *flow* $f :: E \to \mathbb{R}_{\geq 0}$ in Δ satisfies

1. (Capacity restriction) $f(x, y) \leq c(x, y)$ for all $(x, y) \in E$, and
2. (Kirchhoff's 1st law) $d_f^-(x) = d_f^+(x)$ for all $x \in V - \{ s, t \}$.

The *value* $|f|$ of a flow f is f's out-degree of s: $|f| = d_f^+(s)$.

Definition 4 (Orthogonal cut) In a network $\Delta = (V, E, s, t, c)$, a set of vertices C is a *cut* iff $s \in C$ and $t \notin C$. A cut C is *orthogonal* to a flow f iff f saturates all edges going out of C (i.e., $f(x, y) = c(x, y)$ for all $(x, y) \in E$ with $x \in C$ and $y \notin C$) and f is zero on all edges entering C (i.e., $f(x, y) = 0$ for all $(x, y) \in E$ with $x \notin C$ and $y \in C$).

We have already seen an orthogonal pair of a flow of infinite value and a cut in Fig. 1. Another example of an orthogonal flow-cut pair of value 9 is shown in Fig. 2 on the left.

A network constrains the capacities of the edges in a graph, but the throughput of a vertex is unconstrained. So the sums on the two sides of Kirchhoff's first law may be infinite. To avoid such infinite sums, a web constrains the throughput of a vertex and leaves the edge capacity unconstrained. Section 3.1 explains how to convert between networks and webs.

Definition 5 (Web) A *web* $\Gamma = (V, E, A, B, w)$ is a graph (V, E) with two sets of vertices $A, B \subseteq V$ (the sides A and B) and a weight function $w :: V \to \mathbb{R}_{\geq 0}$. We refer to the components of Γ by V_Γ, E_Γ, A_Γ, B_Γ, and w_Γ.

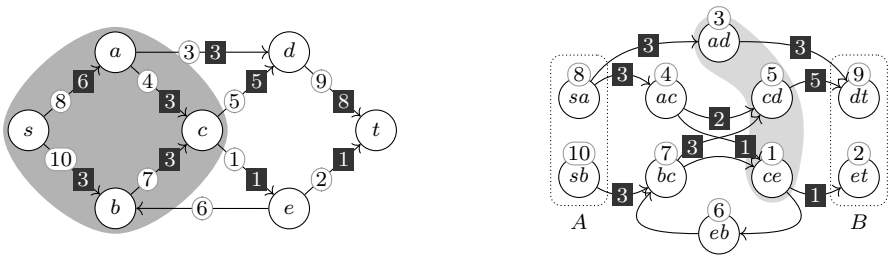

Fig. 3 The network and web from Fig. 2 with a different flow (left) and a web-flow (right).

The two vertex sets A and B correspond to the source and sink of a network, respectively. Currents in a web take the role of flows in a network. The difference is that vertices may leak some of the incoming current (condition 2), i.e., they need not preserve the current.

Definition 6 (Current) Given a web $\Gamma = (V, E, A, B, w)$, a *current* $f :: E \to \mathbb{R}_{\geq 0}$ satisfies

1. (weight restriction) $d_f^-(x) \leq w(x)$ and $d_f^+(x) \leq w(x)$ for all $x \in V$,
2. (flow reflection) $d_f^-(x) \geq d_f^+(x)$ for all $x \in V - A$, and
3. (side restriction) $d_f^-(x) = 0$ for $x \in A$ and $d_f^+(y) = 0$ for $y \in B$.

A current f is called a *web-flow* if $d_f^-(x) = d_f^+(x)$ for all $x \in V - (A \cup B)$. If $d_f^+(x) \geq w(x)$, then f *exhausts* x. If $x \in A$ or $d_f^-(x) \geq w(x)$, then f *saturates* x. A saturated sink x is called *terminal*. The set of saturated vertices is written as $\mathrm{SAT}(f)$ and the set of terminal vertices as $\mathrm{TER}(f) = \mathrm{SAT}(f) \cap \mathrm{SINK}(f)$.

Figure 2 shows an example web on the right where the weight of the vertices are shown in rounded rectangles. It is derived from the network on the left as we will see in Sect. 3.1. The black rectangles specify a current f whose terminal vertices $\mathrm{TER}(f)$ are shown in grey. It exhausts none of the vertices. The current f is not a web-flow because some vertices are leaking, e.g., $d_f^-(bc) = 7 > 6 = d_f^+(bc)$.

Figure 3 shows a different flow and current for same network and web, respectively. The flow on the left differs from the one in Fig. 2 only in that three units are routed through (s, a) and (a, c) instead of through (s, b) and (b, c). So the vertex c now mixes the units coming from a with the three units coming from b and outputs five of them to d and one to e. On the right, a web-flow is shown, which refines the flow on the left as will be explained in Sect. 3.1. The light-grey area contains the exhausted vertices, namely ad, cd, and ce. There are no terminal vertices as the three sinks dt, et, and eb are disjoint from the saturated vertices sa, sb, ad, cd, and ce.

Definition 7 (Essential vertex) Given sets of vertices S and B in a graph $G = (V, E)$, a vertex $x \in S$ is *essential* in S iff there is a path from x to a vertex in B which does not contain a vertex in $S - \{x\}$. The set of essential vertices of S is written as $\mathcal{E}_{G,B}(S)$.

Definition 8 (Separation and roofing) A set S of vertices in graph G *separates* a vertex x from a set of vertices B iff every path from x to a vertex in B contains a

vertex in S. The set S is said to *separate* a set of vertices A from B iff it separates every vertex in A from B.

The *roofing* of S and B (notation $\mathrm{RF}_{G,B}(S)$) consists of all vertices which S separates from B. The *strict roofing* excludes essential vertices: $\mathrm{RF}^\circ_{G,B}(S) = \mathrm{RF}_{G,B}(S) - \mathcal{E}_{G,B}(S)$.

In a web $\Gamma = (V, E, A, B, w)$, S is *A-B-separating* iff it separates A and B. If f is a current in Γ, we abbreviate $\mathcal{E}(f) = \mathcal{E}_{\Gamma,B}(\mathrm{TER}(f))$ and $\mathrm{RF}(f) = \mathrm{RF}_{\Gamma,B}(\mathrm{TER}(f))$ and $\mathrm{RF}^\circ(f) = \mathrm{RF}^\circ_{\Gamma,B}(\mathrm{TER}(f))$.

In the web in Fig. 2, the grey vertices $\mathrm{TER}(f)$ separate A from B. The vertex ac is not essential in $\mathrm{TER}(f)$ as all paths from ac to B pass either through cd or ce, which are both in $\mathrm{TER}(f)$. The roofing $\mathrm{RF}(f)$ contains all the vertices to the left of ad, cd, and ce, inclusive, i.e., $\mathrm{RF}(f) = \{sa, sb, ac, bc, ad, eb, cd, ce\}$. The strict roofing $\mathrm{RF}^\circ(f)$ excludes the essential vertices ad, eb, and ce. Since ac is not essential in $\mathrm{TER}(f)$, the strict roofing includes ac.

Lemma 1 ([2, Lemma 2.14]) *If S separates A from B in G, so does $\mathcal{E}_{G,B}(S)$.*

The key tool for the proof is the concept of a wave. Waves are currents whose terminal vertices separate A from B and which are zero outside of the roofing of the terminal vertices. Intuitively, a wave's essential terminal vertices identify a bottleneck in the web: since the wave saturates them, all other separating sets between the A side and the terminal vertices must allow at least the same current.

Definition 9 (Wave) A current f in Γ is a *wave* iff $\mathrm{TER}(f)$ is A-B-separating and $d^+_f(x) = 0$ for $x \notin \mathrm{RF}(f)$.

In Fig. 2, the current f is 0 outside of $\mathrm{RF}(f)$, i.e., on the edges entering B. So f is a wave. Conversely, the web-flow g in Fig. 3 is not a wave as $\mathrm{TER}(g) = \{\}$ does not separate A from B.

3 From Networks to Bipartite Webs and Back

Aharoni et al.'s proof proceeds in four steps [3]:

1. Transform the network into a web.
2. Find a maximal wave in the web. Its roofing determines the cut.
3. Trim the wave, i.e., reduce the wave such that strictly roofed vertices preserve the current.
4. Extend the wave to a web-flow. This uses a reduction to bipartite webs in which every current is a web-flow by definition.

In this section, we cover these steps up to the reduction to bipartite webs. The next section takes care of actually finding a suitable current in the bipartite web.

3.1 From Networks to Webs

The first step reduces a network Δ to a web, which we denote by $\mathrm{web}(\Delta)$. Every edge e becomes a vertex of $\mathrm{web}(\Delta)$ with weight $c(e)$. Every two incident edges (x, y) and (y, z) in the network induce an edge between the vertices (x, y) and (y, z) in $\mathrm{web}(\Delta)$. The side A consists of the edges leaving s and B of the edges entering t. Formally:

$V_{\text{web}(\Delta)} = E_\Delta$ $\quad E_{\text{web}(\Delta)} = \{((x,y),(y,z)) \mid (x,y) \in E_\Delta \wedge (y,z) \in E_\Delta\}$
$A_{\text{web}(\Delta)} = \{(s,y) \mid (s,y) \in E_\Delta\}$ $\quad B_{\text{web}(\Delta)} = \{(x,t) \mid (x,t) \in E_\Delta\}$
$w_{\text{web}(\Delta)}(e) = c(e)$

For example, Figs. 2 and 3 show the same network Δ on the left and the corresponding web web(Δ) on the right. Webs have the advantage over networks that the current makes explicit how the incoming flow is split up into the outgoing edges of a vertex. In Fig. 3, e.g., the web-flow on the right specifies that the three units flowing from *sa* to *ac* split up into two units going to *cd* and one unit going to *ce*. The flow in the network on the left cannot express this detail: the vertex *c* mixes the two incoming flows of 3 units each and distributes somehow into five and one outgoing units.

Webs therefore allow us to capture flow preservation more precisely than networks. For if a flow f through a network vertex x is infinite, then flow preservation at x merely states that both sums are infinite: $d_f^-(x) = d_f^+(x) = \infty$. This creates problems if we want to subtract two infinite flows f and g from one another because $d_f^-(x) - d_g^-(x) = \infty - \infty$ is not meaningful. So even if both f and g satisfy Kirchhoff's first law at a vertex, it is not clear that their difference $f - g$ satisfies it. In the corresponding web, in contrast, a web-flow g specifies precisely the finite amount each incoming edge contributes to each outgoing edge. So for a web-flow or current g, the sums $d_g^-(x)$ and $d_g^+(x)$ are finite because they are bounded by the finite vertex weights, i.e., the edge capacities in the network. Accordingly, subtraction of flows has nice algebraic properties such as $d_f^-(x) - d_g^-(x) = d_{f-g}^-(x)$ if $f \geq g$.

We next transfer the orthogonality notion from networks to webs. We show that an A-B-separating set S and an orthogonal web-flow f in web(Δ) induce a cut $\hat{S}$ and an orthogonal flow $\hat{f}$ in the original network Δ. Figure 3 illustrates the reduction: The flow $\hat{f}$ in the network Δ on the left corresponds to the web-flow f in web(Δ) on the right. The set $\mathcal{E}(\text{SAT}(f))$ in grey on the right is orthogonal to the web-flow f and yields the cut $\hat{S}$ on the left.

Definition 10 (Orthogonal current) Let $\Gamma = (V, E, A, B, w)$ be a web. A set of vertices S is *orthogonal* to a current f iff

(i) $d_f^-(x) = w(x)$ for $x \in S - A$,
(ii) $d_f^+(x) = w(x)$ for $x \in (S \cap A) - B$, and
(iii) $f(x,y) = 0$ for $x \in V - \text{RF}^\circ(S)$ and $y \in \text{RF}(S)$.

Intuitively, an orthogonal current exhausts the vertices in S unless the vertex belongs to both sides. Condition (iii) ensures that nothing flows back into the roofed vertices. For example, the web-flow in Fig. 4 is not orthogonal to the vertices in the grey area, because one unit flows from the essential vertex *ce* back to the roofed vertex *eb*.

Lemma 2 (Reduction from networks to webs) *Let $\Delta = (V, E, s, t, c)$ be a network with $s \neq t$ and no outgoing edge from t and no direct edge from s to t. Suppose that all edges have positive capacity, i.e., $c(e) > 0$ for $e \in E$.*

(a) Let f be a web-flow in web(Δ). *Define $\hat{f}$ by $\hat{f}(e) = \max(d_f^+(e), d_f^-(e))$ for $e \in E$. Then, $\hat{f}$ is a flow in Δ.*

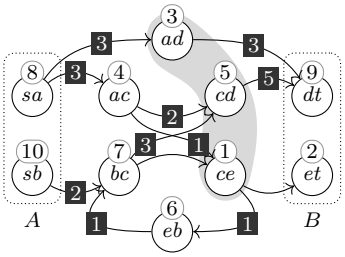

Fig. 4 A separating set (grey area) that is not orthogonal to the shown web-flow.

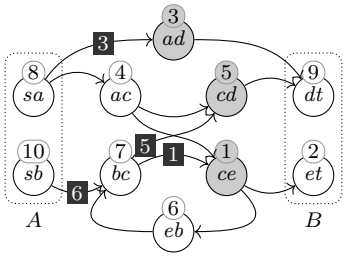

Fig. 5 A trimming of the wave from Fig. 2.

(b) Let S be an A-B-separating set in $\mathrm{web}(\Delta)$*. Define* $\hat{S} = \mathrm{RF}_{\Delta,\{t\}}(\{x \mid \exists y.\, (x,y) \in \mathcal{E}(S)\})$*. Then $\hat{S}$ is a cut in Δ.*
(c) Let an A-B-separating set S be orthogonal to a web-flow f. Then $\hat{S}$ is orthogonal to $\hat{f}$.

By this lemma, to find a cut and an orthogonal flow in a network Δ, it suffices to find a separating set of vertices in $\mathrm{web}(\Delta)$ and an orthogonal web-flow f. In the next section, we focus on finding a suitable separating set, namely the terminal vertices of a maximal wave.

3.2 Maximal Waves and Trimmings

Waves and currents can be ordered pointwise: if f and g are waves or currents in $\Gamma = (V, E, A, B, w)$, then $f \leq g$ iff $f(e) \leq g(e)$ for all $e \in E$. The waves in a countable web form a chain-complete partial order (ccpo), and so do the currents. Therefore, every countable web contains a maximal wave [3, Cor. 4.4] by Zorn's lemma.

Recall that a wave's terminal vertices describe a bottleneck in the web. Intuitively, the maximal wave identifies a narrowest bottleneck in the web: Roughly speaking, the roofed part cannot contain a tighter bottleneck because if so, the current could not saturate the terminal vertices due to the flow reflection condition. Conversely, if a separating set beyond the terminal vertices formed a tighter bottleneck, then we could extend the wave and saturate that smaller bottleneck, which contradicts maximality. Here, it is crucial that a wave may partially leak the incoming current of some vertices, i.e., they need not preserve the current.

A trimming of a wave reduces the current such that the incoming current is preserved on the strict roofing. For example, the wave in Fig. 2 on the right is maximal. Its trimming is shown in Fig. 5. The current is reduced on the edge from *sb* to *bc* from 7 to 6 and on the edge from *sa* to *ac* from 4 to 0.

Definition 11 (Trimming) Let f be a wave in $\Gamma = (V, E, A, B, w)$. A wave g is called a *trimming* of f iff

(i) $g \leq f$,
(ii) $d_g^+(x) = d_g^-(x)$ for all $x \in \mathrm{RF}^\circ(f) - A$, and
(iii) $\mathcal{E}(\mathrm{TER}(g)) - A = \mathcal{E}(\mathrm{TER}(f)) - A$.

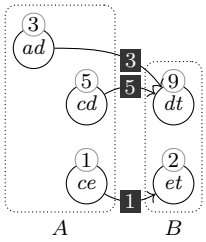

Fig. 6 The quotient of the web and wave of Fig. 2 with a linkage.

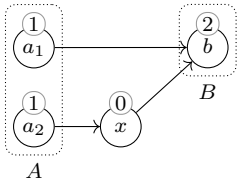

Fig. 7 A web that contains no non-zero wave, but the zero wave is a hindrance.

Lemma 3 ([3, Lemma 4.8]) *Every wave in a countable web has a trimming.*

Proof The trimming for a wave f is constructed as the transfinite fixpoint iteration of the one-step trimming function $trim_1$ starting at f. For a wave g, $trim_1(g)$ picks some strictly roofed vertex z where Kirchhoff's first law does not hold, i.e., $z \in \mathrm{RF}^\circ(g) - A \wedge d_g^+(z) \neq d_g^-(z)$. Then, $trim_1$ reduces the current on z's incoming edges by the factor $\frac{d_g^+(z)}{d_g^-(z)}$ so that Kirchhoff's first law holds at z afterwards.

$$trim_1(g)(y,x) = \begin{cases} g(y,x) & \text{if } g \text{ is a trimming} \\ \mathsf{if}\ x = z\ \mathsf{then}\ \frac{d_g^+(z)}{d_g^-(z)} * g(y,x)\ \mathsf{else}\ g(y,x) & \text{if such a } z \text{ exists} \end{cases}$$

The fixpoint exists by Bourbaki-Witt's fixpoint theorem [12] as $trim_1$ is decreasing, i.e., $trim_1(g) \leq g$, and the set of waves g with $g \leq f$ is a chain-complete partial order w.r.t. $\geq$. The proof that the fixpoint satisfies the trimming conditions relies on d^+ and d^- being point-wise order-continuous, which holds by monotone convergence as the web is countable. □

3.3 A Linkage in the Quotient of a Web

The trimming of a maximal wave f describes the first half of the web-flow we are looking for (Fig. 5). For the second half, we consider the residual web beyond f's terminal vertices, which is called the quotient Γ/f. Figure 6 shows the quotient for the web and wave f from Fig. 2. The essential terminal vertices of the wave become the side A. The quotient does not include the roofed vertex eb even though it is reachable from $\mathcal{E}(\mathrm{TER}(f))$ as we want to construct an orthogonal current and nothing may flow back into roofed vertices. The formal definition is a bit complicated so that it also works when there are edges between vertices in $\mathcal{E}(\mathrm{TER}(f))$ or when $\mathcal{E}(\mathrm{TER}(f))$ contains vertices from B. The details are discussed in Sect. 6.

Definition 12 (Quotient) Let $\Gamma = (V, E, A, B, w)$ and f be a wave in Γ. The quotient Γ/f is the following web:

$$\begin{aligned} V_{\Gamma/f} &= V_\Gamma - (\mathrm{RF}^\circ_\Gamma(f) \cup (\mathrm{TER}_\Gamma(f) \cap B_\Gamma)) \\ E_{\Gamma/f} &= \{(x,y) \in E_\Gamma \mid x \notin \mathrm{RF}^\circ_\Gamma(f) \wedge y \notin \mathrm{RF}_\Gamma(f)\} \\ A_{\Gamma/f} &= \mathcal{E}_\Gamma(\mathrm{TER}_\Gamma(f)) - (B_\Gamma - A_\Gamma) \\ B_{\Gamma/f} &= B_\Gamma \\ w_{\Gamma/f}(x) &= \begin{cases} w(x) & \text{for } x \in V_\Gamma - (\mathrm{RF}^\circ_\Gamma(f) \cup (\mathrm{TER}_\Gamma(f) \cap B_\Gamma)) \\ 0 & \text{for } x \in \mathrm{TER}_\Gamma(f) \cap B_\Gamma \end{cases}. \end{aligned}$$

In the quotient Γ/f, we now look for a web-flow g that saturates all vertices in A, i.e., $\mathrm{TER}(f)$. Such a web-flow is called a linkage. Then, the web-flow in Γ is given by the trimming of f plus g. Figure 6 shows such a linkage; together with the trimmed wave from Fig. 5, they form the orthogonal web-flow whose reduction (Lemma 2) yields the network flow shown in Fig. 2.

Definition 13 (Linkage [3, Def. 4.1]) A web-flow f in a web $\Gamma = (V, E, A, B, w)$ is called a *linkage* iff f exhausts all vertices in A, i.e., $d_f^+(a) = w(a)$ for all $a \in A$.

Under what conditions does a web Γ contain a linkage? Certainly, there must not be a bottleneck beyond the A side. Waves describe such bottlenecks. So if the zero wave is the only wave in Γ, then the A side is the only bottleneck. Moreover, we need that all vertices in A are essential for separation unless their weight is 0. For example, the web in Fig. 7 contains only the zero wave, but not a linkage. The problem is that the vertex a_2 with weight 1 is bottlenecked by the zero-weight vertex $x \in \mathcal{E}(\mathrm{TER}(\mathbf{0}))$. Such a situation is called a hindrance.

Definition 14 (Hindrance, looseness, [3, Def. 4.5]) A wave f in a web $\Gamma = (V, E, A, B, w)$ is a *$>\varepsilon$-hindrance* iff there is a vertex $a \in A - \mathcal{E}(\mathrm{TER}(f))$ such that $\varepsilon < w(a) - d_f^+(a)$. Also, f is a *hindrance* iff there exists a $\varepsilon > 0$ such that f is a $>\varepsilon$-hindrance. A web is called *hindered* (respectively *$>\varepsilon$-hindered*) iff it contains a hindrance (respectively a $>\varepsilon$-hindrance). A web is called *loose* iff it contains no non-zero wave and the zero wave is not a hindrance.

Lemma 4 ([3]) *If f is a maximal wave in the web Γ, then Γ/f is loose.*

3.4 Reduction to Bipartite Webs

To find linkages in countable loose webs, Aharoni et al. [3] transform webs into bipartite webs. A web $\Omega = (V, E, A, B, w)$ is *bipartite* iff there are only edges from nodes in A to nodes in B, i.e., iff $V = A \cup B$ and $A \cap B = \emptyset$ and $E \subseteq A \times B$.

We briefly review the transformation described in [1]; Fig. 8 shows an example. In this section, we always assume that the web $\Gamma = (V, E, A, B, w)$ has no incoming edges to vertices in A, no outgoing edges from vertices in B, no loops, and that A and B are disjoint. In the bipartite web $\mathrm{bp}(\Gamma)$, there are two copies x' and x'' for every vertex $x \in V - (A \cup B)$. Vertices $x \in A$ and $y \in B$ only have one copy x' and y'', respectively. The edges are $E_{\mathrm{bp}(\Gamma)} = \{(x', y'') \mid (x, y) \in E\} \cup \{(x', x'') \mid x \in V - (A \cup B)\}$ and the sides $A_{\mathrm{bp}(\Gamma)} = \{x' \mid x \in V - B\}$ and $B_{\mathrm{bp}(\Gamma)} = \{x'' \mid x \in V - A\}$ and the weight function $w(x') = w(x)$ for $x \in V - B$ and $w(x'') = w(x)$ for $x \in V - A$.

An A-B-separating set S in $\mathrm{bp}(\Gamma)$ induces an A-B-separating set $\widetilde{S}$ in Γ given by $\widetilde{S} = (A_S \cap B_S) \cup (A \cap A_S) \cup (B \cap B_S)$ where $A_S = \{v \mid v' \in S\}$ and $B_S = \{v \mid v'' \in S\}$ [1]. Moreover, a wave f in $\mathrm{bp}(\Gamma)$ induces a wave $\tilde{f}$ in Γ given by $\tilde{f}(x, y) = f(x', y'')$ for $(x, y) \in E$ with $\mathrm{TER}_\Gamma(\tilde{f}) = \widetilde{\mathrm{TER}_{\mathrm{bp}(\Gamma)}(f)}$ [3, Lemma 6.3].

Lemma 5 *If Γ is loose, then $\mathrm{bp}(\Gamma)$ is unhindered.*

Aharoni et al. wrongly claimed the stronger statement that if Γ is loose then $\mathrm{bp}(\Gamma)$ is loose [3, below Thm. 6.5]. We provide a counterexample in Sect. 6. Note that the reduction bp does not preserve unhinderedness either.

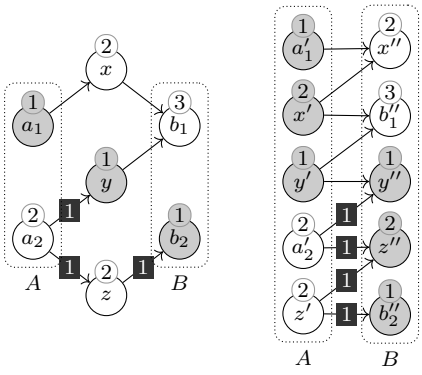

Fig. 8 An unhindered web Γ (left) and its bipartite reduction $\mathrm{bp}(\Gamma)$ (right). The wave f in $\mathrm{bp}(\Gamma)$ induces the wave $\tilde{f}$ in Γ.

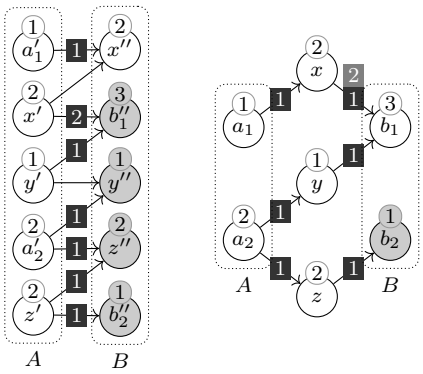

Fig. 9 A linkage g in $\mathrm{bp}(\Gamma)$ (left) that yields a linkage (right) in the web Γ from Fig. 8 by trimming $\tilde{g}$ at vertex x.

Conversely, a linkage g in $\mathrm{bp}(\Gamma)$ yields a linkage in Γ as illustrated in Fig. 9: For $\tilde{g}$ as defined above, we have $d^+_{\tilde{g}}(a) = d^+_g(a') = w(a)$ for $a \in A_\Gamma$ and $d^+_{\tilde{g}}(x) \geq d^-_{\tilde{g}}(x)$ for all $x \notin B$. So the out-flow of some vertices may surpass the in-flow, e.g., x in Fig. 9. Analogously to the trimming of waves, we can trim $\tilde{g}$ using a fixpoint iteration to obtain the linkage in Γ.

Lemma 6 ([3]) *If* $\mathrm{bp}(\Gamma)$ *contains a linkage and* Γ *is countable, then* Γ *contains a linkage.*

4 Linkability in unhindered bipartite webs

By the results in Sect. 3, the max-flow min-cut theorem for the countable case (Thm. 1) follows from the following theorem, which we prove in this section.

Theorem 2 (Bipartite linkability) *A countable unhindered bipartite web contains a linkage.*

In fact, we present two ways how to construct such a linkage in an unhindered bipartite web. Both ways enumerate the vertices in $A = \{a_1, a_2, a_3, \ldots\}$ and construct a sequence of web-flows f_i that exhaust $\{a_1, \ldots, a_i\}$ so that the limit f exhausts all of A. The difference is in how the f_i are constructed and in the limit argument. In Sect. 4.1, each f_i is constructed independently as the limit of maximum flows in a finite network; the existence and the linkage property of the limit for these f_i themselves is shown using diagonalization and majorised convergence. Unfortunately, this construction only works if the neighbours of any a_i vertex have finite total weight.

In contrast, f_{i+1} in Sect. 4.2 saturates a_{i+1} by extending the previous web-flow f_i with a sequence of augmenting flows in the so-called residual network, similar to how classic max-flow algorithms for finite networks work [15]. This construction avoids taking infinite summations and thus yields a proof of Thm. 2 without additional assumptions. However, the proof is more involved than in the bounded case.

4.1 The Bounded Case

We first prove Thm. 2 for the case where the neighbours of each vertex in A have only bounded total weight, i.e., $\sum_{y \in \mathrm{OUT}(x)} w(y) < \infty$ for all $x \in A$. The general case is shown in the next section.

The next lemma states the crucial property of unhindered bipartite webs, namely that the total weight of any finite set of A vertices is at most the total weight of their neighbours in B.

Lemma 7 *Let $\Omega = (V, E, A, B, w)$ be a countable unhindered bipartite web and $X \subseteq A$ be finite. Then, $\sum_{x \in X} w(x) \leq \sum_{y \in E[X]} w(y)$ where $E[X] = \{y \mid \exists x \in X.\ (x, y) \in E\}$ denotes the neighbours of X.*

This lemma allows us to understand a linkage in an unhindered bipartite web as an $A \times B$ matrix over the reals where the weights on A are the row sums of the countable matrix and the edges describe the matrix elements that may be non-zero. In the proof below, we will use the following result about the existence of a countable matrix with given marginals.

Proposition 1 (Matrix with given marginals) *Let $f : A \to \mathbb{R}_{\geq 0}$ and $g : B \to \mathbb{R}_{\geq 0}$ for countable sets A, B such that $\sum_{i \in A} f(i) = \sum_{j \in B} g(j) < \infty$, and let $R \subseteq A \times B$. Assume that $\sum_{i \in X} f(x) \leq \sum_{j \in R[X]} g(j)$ for all $X \subseteq A$. Then, there exists a function $h : A \times B \to \mathbb{R}_{\geq 0}$ such that for all $i \in A$ and $j \in B$:*

- *$h(i, j) = 0$ if $(i, j) \notin R$,*
- *$f(i) = \sum_{j \in B} h(i, j)$, and*
- *$g(j) = \sum_{i \in A} h(i, j)$.*

Proposition 1 follows easily from the following proposition by Kellerer, which is an instance of Strassen's theorem [35]. We have formalized neither Kellerer's proposition nor Strassen's theorem; instead, we adapted Kellerer's proof so that we directly prove Prop. 1. This proof uses the max-flow min-cut theorem for *finite* networks.

Proposition 2 ([20, Satz 4.1]) *Let $f, g : \mathbb{N} \to \mathbb{R}_{\geq 0}$ and $t : \mathbb{N} \times \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that $\sum_{j \in \mathbb{N}} t(i, j) < \infty$ for all $i \in \mathbb{N}$, and $\sum_{i \in \mathbb{N}} t(i, j) < \infty$ for all $j \in \mathbb{N}$, and for all sets $X, Y \subseteq \mathbb{N}$,*

$$\sum_{i \in X} f(i) \leq \sum_{\substack{i \in X \\ j \in Y}} t(i, j) + \sum_{j \in \mathbb{N} - Y} g(j) \qquad \text{and} \qquad \sum_{j \in Y} g(j) \leq \sum_{\substack{i \in X \\ j \in Y}} t(i, j) + \sum_{i \in \mathbb{N} - X} f(i)$$

Then, there exists a function $h : \mathbb{N} \times \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that $h \leq t$ and $f(i) = \sum_{j \in \mathbb{N}} h(i, j)$ and $g(j) = \sum_{i \in \mathbb{N}} h(i, j)$ for all $i, j \in \mathbb{N}$.

We can now prove bipartite linkability in the bounded case. The proof starts with a sequence of increasing finite subsets A_n of A that converge to A, and suitable, possibly infinite subsets B_n of their neighbours in B. For these subsets, we obtain an $A_n \times B_n$ matrix h_n with the right marginals. This sequence h_n converges and its limit yields the desired linkage, using a majorised convergence argument with the bound on the neighbours.

Theorem 3 (Bounded bipartite linkability) *A countable unhindered bipartite web $\Omega = (V, E, A, B, w)$ contains a linkage if $\sum_{y \in \mathrm{OUT}(x)} w(y) < \infty$ for all $x \in A$.*

Together with the reduction from Sect. 3, this yields a proof for Thm. 1 when only the source s and the sink t in the network $\Delta = (V, E, s, t, c)$ may have outgoing edges whose total capacity is infinite, i.e., $d_c^+(x) < \infty$ for $x \in V - \{s, t\}$. The max-flow min-cut use cases in probability theory [33] and privacy [9] satisfy this condition, as we show in Sect. 7.

4.2 The Unbounded Case

We now show that Thm. 2 holds even when the neighbours of a vertex have infinite total weight. Our proof generalizes Aharoni et al.'s from loose to unhindered bipartite webs. For the remainder of this section, we always assume that $\Omega = (V, E, A, B, w)$ is a countable bipartite web. We write $\Omega \ominus f$ for the bipartite web Ω where the weight of the vertices has been reduced by the current f that flows through them.

Definition 15 (Residual web) If $\Omega = (V, E, A, B, w)$ is a bipartite web and f a current in Ω, we write $\Omega \ominus f$ for the web (V, E, A, B, w') where the new weight function w' is given by $w'(x) = w(x) - d_f^+(x)$ for $x \in A$ and $w'(x) = w(x) - d_f^-(x)$ for $x \in B$.

The proof rests on the following step: If Ω is unhindered, then we can find a current f that saturates some vertex $a \in A$ such that the residual web $\Omega \ominus f$ is unhindered again.

Lemma 8 (Vertex saturation in unhindered bipartite webs) *If Ω is unhindered and $a \in A$, then there exists a current f in Ω such that $d_f^+(a) = w(a)$ and $\Omega \ominus f$ is unhindered.*

With this lemma, we can now prove that countable unhindered bipartite webs are linkable (Thm. 2). The proof is analogous to [3, Thm. 6.5], but uses our Lemma 8 instead.

Proof (Proof of Thm. 2) Enumerate the vertices in A as $a_1, a_2, \ldots$. Recursively define a family f_n of currents in Ω as follows:

(i) f_0 is the zero current.
(ii) For $n > 0$, pick a current g_n in $\Omega \ominus f_{n-1}$ such that $d_g^+(a_n) = w_{\Omega \ominus f_{n-1}}(a_n)$ and $\Omega \ominus f_{n-1} \ominus g$ is unhindered. Set $f_n = f_{n-1} + g$.

A simple induction on n shows that f_n is a well-defined current in Ω and $\Omega \ominus f_n$ is unhindered for all n; here, Lemma 8 applied to $\Omega \ominus f_{n-1}$ ensures that g_n exists. Set $g(e) = \sup\{f_n(e) \mid n \in \mathbb{N}\}$ for $e \in E$. Then, g is a current in Ω with $d_g^+(x) = w(x)$ for all $x \in A$. As every current in a bipartite web is a web-flow, g is the linkage we are looking for. □

The proof of the saturation lemma 8 uses the following theorems and lemmas, which have already been proven by Aharoni et al. [3]. We have formalized all of them and fixed the glitches in the original statements and proofs.

Theorem 4 (Flow attainability [3, Thm. 5.1]) *Let $\Delta = (V, E, s, t, c)$ be a countable network with $s \neq t$, no loops and no incoming edges to s, and such that for all $x \in V - \{t\}$, the sum of capacities of the incoming edges to x or the sum of capacities of the outgoing edges from x is finite, i.e., $d_c^-(x) < \infty$ or $d_c^+(x) < \infty$. Then there exists a flow f in Δ such that $d_f^+(s) = \sup\{|g| \mid g \text{ is a flow in } \Delta\}$ and $d_f^-(x) \leq |f|$ for all $x \in V$.*

Lemma 9 ([3, Lemma 6.7]) *Let $\Omega = (V, E, A, B, w)$ be a countable bipartite web and let $u :: V \to \mathbb{R}_{\geq 0}$ such that $u(x) = 0$ for $x \in A$, $u(y) \leq w(y)$ for $y \in B$, and $\varepsilon = \sum_{x \in B} u(x) < \infty$. Let $\Omega' = (V, E, A, B, w - u)$ be the web Ω with w reduced by u. If Ω' is $>\varepsilon$-hindered, then Ω is hindered.*

Lemma 10 ([3, Cor. 6.8]) *Let g be a current in Ω with $\varepsilon := \sum_{b \in B} d_g^-(b) < \infty$. If $\Omega \ominus g$ is $>\varepsilon$-hindered, then Ω is hindered.*

Lemma 11 ([3, Lem 6.9]) *Let Ω be loose and $b \in B$ with $w(b) > 0$. For every $\delta > 0$, there exists an $\varepsilon > 0$ such that $\varepsilon < \delta$ and Ω with the weight of b reduced by ε is unhindered.*

5 Discussion of the Formalization

We have formalized all definitions, theorems, and proofs mentioned in this paper in Isabelle/HOL. This includes all the lemmas and underlying theory; informal proofs can be found in the accompanying report [29]. In this section, we discuss the challenges we faced and the design decisions we made. The issues with the original definitions, theorems, and proofs and their corrections are discussed in the next section.

Graphs are formalized using Isabelle's record package [31] as an extensible record with one field for the edge relation, given as a binary predicate over the vertices of type α. This yields the projection function $\mathsf{edge} :: \alpha\ \mathsf{graph} \Rightarrow \alpha \Rightarrow \alpha \Rightarrow \mathsf{bool}$ for the edge field.[1] From this, we derive the set E of edges as an abbreviation.

record α graph = edge :: $\alpha \Rightarrow \alpha \Rightarrow$ bool
definition vertex :: α graph $\Rightarrow \alpha \Rightarrow$ bool
  where vertex $G\ x = (\exists y.\ \mathsf{edge}\ G\ x\ y \vee \mathsf{edge}\ G\ y\ x)$
type-synonym α edge = $\alpha \times \alpha$
abbreviation E :: α graph $\Rightarrow \alpha$ edge set where $\mathsf{E}_G = \{(x, y).\ \mathsf{edge}\ G\ x\ y\}$

We derive the set of vertices from edges of the graph rather than modelling them separately. This has the advantage that we encode the condition $E \subseteq V \times V$ in the construction and do not have to carry around this well-formedness condition in our formalization. Conversely, graphs in this model cannot have isolated vertices. This is without loss of generality as isolated vertices cannot contribute to any flow or to the edges entering or leaving a cut.

Networks are formalized as an extension of the record graph. So all operations on graphs also work for networks. The same applies to webs.

[1] The record package achieves extensibility with structural subtyping by internally generalizing α graph to (α, β) graph-scheme, where β is the extension slot for further fields. For example, β is instantiated with the singleton type unit for graph. All operations on graph are actually defined on graph-scheme so that they also work for all record extensions. We omit this technicality from the presentation.

record α network = α graph +
  capacity :: $\alpha \Rightarrow$ ennreal
  source :: α
  sink :: α

record α web = α graph +
  weight :: $\alpha \Rightarrow$ ennreal
  A :: α set
  B :: α set

Records provide a simple and lightweight means for grouping the components of a network or web. Particular properties such as countability, finite capacity and weights, and disjoint sides A and B, are formalized as locales [7]. For example, the locale countable-network below enforces that there are only countably many edges, the source is not the sink, and the capacities are finite and 0 outside of the edges. Using the **(structure)** annotation on a record variable like Δ [6], we can omit the network (or web) as subscripts, e.g., in the assumption countable E; Isabelle automatically fills in the corresponding parameter. We use this notational convenience mainly for definitions that need custom syntax anyway, e.g., $\mathcal{E}$, RF, and RF$^\circ$. For plain HOL functions without special syntax like capacity and source, it is usually faster to type the record parameter than to enter special syntax.

locale countable-network = **fixes** Δ :: α network **(structure)**
  assumes countable E **and** source $\Delta \neq$ sink Δ
    and $e \notin$ E $\Longrightarrow$ capacity Δ $e = 0$ **and** capacity Δ $e < \infty$

Since flows, cuts, and capacities are always non-negative, we use the extended non-negative reals ennreal from Isabelle/HOL's library everywhere. Summations like the in-degree d^- are expressed using the Lebesgue integral nn-integral over the counting measure count-space A on the set A. So every subset of A is measurable and all points have equal weight. Moreover, every function is integrable and we need not discharge neither integrability nor summability conditions in the proofs. Just the finiteness conditions of the form $\sum_{x \in A} \ldots < \infty$ are ubiquitous.

We also formalize capacities and weights as ennreal and explicitly require them being finite in the locales. This avoids coercions from the real numbers real into ennreal, which would complicate the proof formalization. For example, the in-degree $d_f^-(f)$ of y is defined as

definition d-IN :: (α edge $\Rightarrow$ ennreal) $\Rightarrow$ $\alpha \Rightarrow$ ennreal
  where d-IN f $y = \sum_{x \in \mathsf{UNIV}} f\ (x, y)$

where $\sum_{x \in A} g$ desugars to nn-integral (count-space A) ($\lambda x.\ g$). We let the summation range over UNIV, the set of all values of α, not only the neighbours of y. Instead, we enforce that f is 0 outside of E, e.g., via the capacity assumption in countable-network. This way, d-IN depends only on f and not on the graph. This simplifies the formalization because when we consider f in the context of different graphs, d-IN f is trivially the same for all of them.

Regarding the mathematical background theory, we found that most relevant theorems were readily available in the Isabelle/HOL library: limits, infinite summations via the Lebesgue integral, monotone and majorised convergence, lim sup and lim inf. There is even a generic formalization of Cantor's diagonalization argument by Immler [19]. The Bourbaki-Witt fixpoint theorem [12], however, was missing. We therefore ported the Coq formalization by Smolka et al. [34] to Isabelle/HOL. It is now part of Isabelle/HOL's library. We have also contributed many lemmas about ennreal and nn-integral to the library.

Apart from identifying and fixing glitches and mistakes in definitions and proofs (Sect. 6), we faced three main challenges during the formalization. First, the def-

inition and proof principles in the paper are often not suitable for direct formalization. For example, the original proofs construct trimmings, linkages and saturating flows using transfinite iteration and transfinite induction with ordinals. We have replaced them with fixpoints of increasing or decreasing functions in a chain-complete partial order, using Bourbaki-Witt's fixpoint theorem (Lemmas 3, 6, and 8). This way, we did not need to formalize ordinals and their theory.

Second, applying the theorems from the Isabelle library often needs a small twist. The proof for the existence of a maximal wave in Sect. 3.2 demonstrates this. The proof that the least upper bound $\bigsqcup_{i \in I} f_i$ for a chain f_i of currents in a web Γ is a current relies on Beppo Levi's monotone convergence theorem. The challenge here was that the monotone convergence theorem applies only to countable increasing sequences, whereas Isabelle's formalizaton of chain-complete partial orders demands the existence of least upper bounds for arbitrary (uncountable) chains. We bridge the gap by finding a countable subsequence of any such chain, which relies on the currents being non-zero only on the countably many edges.

Third, we often faced the problem that a statement had some precondition that was not met when we wanted to apply it. In an informal proof, these preconditions would be assumed "without loss of generality" or ignored altogether. We deal with them in two ways: either introduce a reduction that ensures the precondition, or generalize the definitions and proofs so that the precondition is not needed. Reductions are in general preferable as generalizations often complicate the definitions and proofs. Additional reductions can be seen, e.g., in Lemma 2. It assumes that there is no direct edge from s to t and all edges have positive capacity. The final theorem 1 does not make these assumptions. We therefore introduce another reduction that splits a potential s-t edge by introducing a new vertex and removes all edges with no capacity. Similarly, the reduction to bipartite webs in Sect. 3.4 assumes that the web does not contain loops. These loops would originate from loops in the original network; so we have another reduction that eliminates loops in networks. Reductions are not always feasible though. The example of the quotient web (Def. 12) is discussed in the next section.

On the positive side, reasoning about paths in networks and webs was much less of a pain than we had expected. We formalized a finite path as a list of vertices, which allows us to reuse Isabelle's library for lists to manipulate and reason about paths. For example, the predicate distinct expresses that a path does not contain cycles, and $\pi \mathbin{@} [x] \mathbin{@} \pi'$ denotes the concatenation of the two paths $\pi \mathbin{@} [x]$ and $[x] \mathbin{@} \pi'$. Moreover, we found that $\mathcal{E}$, RF, and RF° are powerful concepts that allow us to avoid explicitly dealing with paths in the main lemmas about flows—once we had proven enough properties about them.

Table 1 shows line counts of the Isabelle theories for different parts of the formalization, as a proxy for the formalization effort. These counts exclude empty lines. The left part lists the material that is used by both linkability proofs for bipartite webs. This covers the concepts of networks, flows, webs, currents, (maximal) waves, and trimmings, as well as the reductions from networks to webs and from webs to bipartite webs. On the right, the line counts are shown for linkability of bounded (Sect. 4.1) and unbounded (Sect. 4.2) countable bipartite webs, together with the line counts for the helper statements 1 and 4. The unbounded case requires about 3.6 times as much space as the bounded case if we include the formalization of the helper statements. If we exclude the helper statements, the ratio is about 5.4. This highlights how much more complicated the general case is.

Table 1 Line counts for different parts of the formalization, not counting empty lines

	Shared		Bounded	Unbounded
preliminaries	200	matrix for marginals (Prop. 1)	845	
networks & webs	2214	flow attainability (Thm. 4)		1954
reductions	1248	bipartite linkability (Thms. 3 / 2)	593	3158
total	3662		1436	5112

We have also generated a PDF from the Isabelle theories using Isabelle's document preparation system. The material corresponding to shared and unbounded fill 236 pages. Aharoni et al. need a bit more than 10 pages in [3]. This gives an expansion factor of about 23. This is much higher than for text book mathematics, where the factor is typically well below 10 [8, 36]. We take this as an indication that the original paper is very dense.

6 Problems in the Original Proof

We now discuss the problems we have identified in the original paper during the formalization.

Reduction to bipartite webs This is the main problem we have found. Aharoni et al. [3] claim that the reduction to bipartite webs from Sect. 3.4 preserves looseness, but this is not the case. In Fig. 10, the web Γ on the left is loose, its bipartite transformation $\mathrm{bp}(\Gamma)$ on the right is not loose, because it contains the non-zero wave shown. The problem is that there is no path from the (infinitely many) vertices y_i (where $i \in \mathbb{N}$) to b. In a finite web, we could remove all vertices that cannot reach a vertex in B, because they cannot contribute to a web-flow. In the infinite case, however, we cannot do so easily because such infinite paths do occur in infinite networks and absorb parts of the (maximal) flow; an example is given in the conclusion. So their key theorem [3, Thm. 6.5], namely that every countable loose bipartite web contains a linkage, cannot be used to prove the general case.

Instead, we strengthen the theorem to countable *unhindered* bipartite webs (Thm. 2). The induction invariant now is $\Omega \ominus f_n$ being unhindered rather than being loose, and the induction step (Lemma 8) must also be generalized. Fortunately, the original high-level ideas carry over; our proof composes the lemmas 9, 10 and 11 in a different order. We regain looseness from unhinderedness by first finding a maximal wave and reducing the weights, similar to what is happening in Lemma 4. Note that the reduction bp does not preserve unhinderedness either, as the example in Fig. 11 shows. The web on the left is not loose as it contains the shown wave.

Quotient webs Quotient webs (Def. 12) are an example where the definition had to be changed. This change propagates to the proofs of the basic properties of quotient webs. In detail, the original definition sets the edges as $E_{\Gamma/f} = \{(x, y) \in E \mid x \notin \mathrm{RF}^\circ_\Gamma(f) \wedge y \notin \mathrm{RF}^\circ_\Gamma(f)\}$, i.e., an edge may point to one of f's essential terminal vertices. Our Definition 12 excludes these edges. The difference is illustrated in Fig. 12. The quotient Γ/f on the right of the web Γ and the wave f on the left

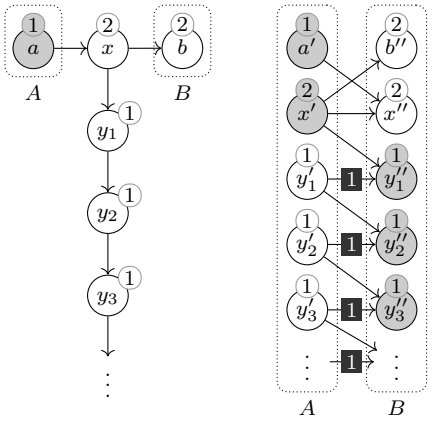

Fig. 10 A loose web (left) whose bipartite reduction (right) is not loose as witnessed by the non-zero wave shown.

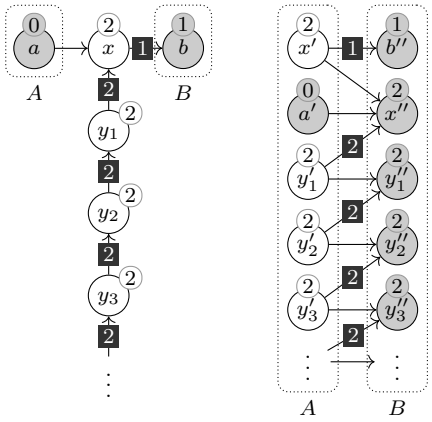

Fig. 11 An unhindered web (left) whose bipartite reduction (right) contains a hindrance as witnessed at x'.

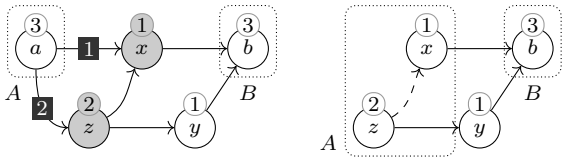

Fig. 12 A wave f in a web Γ (left) and the quotient web Γ/f (right). The quotient contains the problematic edge (z, x) only in [3].

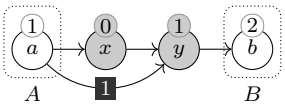

Fig. 13 Wave f in a web none of whose trimmings g satisfies Aharoni et al.'s condition $\mathrm{TER}(g) - A = \mathcal{E}(\mathrm{TER}(f)) - A$.

contains the edge (z, x) only with the original definition. This edge invalidates a number of statements, e.g., that $f + g \restriction (\Gamma/f)$ is a current or a wave if g is a current or a wave in Γ, where $g \restriction (\Gamma/f)$ restricts g to the vertices of Γ/f. Take, e.g., $g(a, z) = 2$, $g(z, x) = g(z, y) = 1$, and $g(e) = 0$ otherwise.

Our definition therefore excludes this edge. And while we were at it, we also changed the definition of $A_{\Gamma/f}$ and the weights so that the two sides of the quotient are always disjoint and vertices without edges have weight 0. These changes ensure that the quotient web meets the assumptions of the reduction to bipartite webs (Sect. 3.4). Accordingly, we had to adapt the existing proofs about the quotient web's properties or find new ones.

Trimmings The definition of trimmings (Def. 11) is an example of a small glitch that affects proofs only minimally. For trimmings, Aharoni et al. [3] require the stronger condition $\mathrm{TER}(g) - A = \mathcal{E}(\mathrm{TER}(f)) - A$ instead of $\mathcal{E}(\mathrm{TER}(g)) - A = \mathcal{E}(\mathrm{TER}(f)) - A$. The two are equivalent only if there are no vertices with weight 0, but webs may contain such vertices. So Lemma 3 need not hold for such webs. For example, Fig. 13 shows a wave f that does not have a trimming according to Aharoni et al.'s definition [3, Def. 4.7]. Every wave g has $x \in \mathrm{TER}(g)$ because x has weight 0, but $x \notin \mathcal{E}(\mathrm{TER}(f)) - A = \{y\}$.

Reduction from networks to webs The first step in the proof reduces networks to web (Sect. 3.1). The reduction in [3] contains two flaws, which we have fixed.

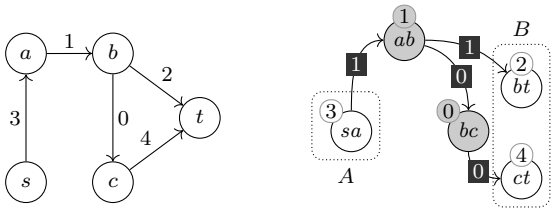

Fig. 14 A network (left) and the corresponding web (right) which contains an A-B-separating set of terminal vertices (grey) which do not correspond to a cut of the network.

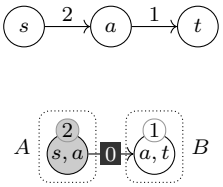

Fig. 15 The network on the top shows that condition (ii) in Def. 10 is needed for the reduction to the web at the bottom.

First, Aharoni et al. [3] define a cut as a set of edges of the form $\{(x, y) \in E \mid x \in S \wedge y \notin S\}$ for some set of vertices S such that $s \in S$ and $t \notin S$. They claim that if S is A-B-separating in web(Δ), then $\mathcal{E}(S)$ is a cut. This need not hold as the example in Fig. 14 shows. The grey web vertices ab and bc separate A and B and are both essential (ab is essential due to the edge to bt and bc due to the edge to ct). But the set $\{(a, b), (b, c)\}$ of corresponding edges in the network is no cut, because b occurs both as the end and as the start of an edge. As the two grey vertices in Fig. 14 are orthogonal to the web-flow, the reduction as stated in [3] fails for this network.

Instead, we define the cut $\hat{S}$ corresponding to an A-B-separating set S as the roofing of the source vertices of the edges in S. Moreover, A-B-separating sets orthogonal to a web-flow can only contain two neighbouring web vertices if one of them has weight 0. Therefore, our Lemma 2(c) requires that all network edges have positive capacity.

Second, the original definition of orthogonality in webs [3] is too permissive. In detail, they call an A-B-separating set S orthogonal to a web-flow f iff $S \subseteq \mathrm{SAT}(f)$ and $f(x, y) = 0$ for all $x \in V - \mathrm{RF}^\circ(f)$ and $y \in \mathrm{RF}^\circ(f)$. Our notion of orthogonality strengthens theirs in two respects. First, we change $y \in \mathrm{RF}^\circ(f)$ to $y \in \mathrm{RF}(f)$. This is necessary to avoid the problem from Fig. 14. Second, we add the condition (ii) in Def. 10. Figure 15 shows why the condition is needed. The grey vertex A-B-separates the web at the bottom and is orthogonal to the zero web-flow. Yet, the edge (s, a) is not orthogonal to the zero flow in the network at the top.

Flow attainability The proof of the unbounded bipartite case (Thm. 2) makes use of the flow attainability theorem (Thm. 4). Aharoni et al. [3, Thm. 5.1] have proved it in the special case when $d_c^-(x) < \infty$ for all $x \in V$ and there are no incoming edges to s. A careful analysis shows that their proof generalises to our statement.

Vertex saturation in bipartite webs The proof of Lem. 8 constructs the desired current by transfinitely iterating a saturation function *sat* over a pair of a current f_n in Ω and a wave h_n in $\Omega \ominus f$ with two invariants:

- all vertices other than the vertex a to be saturated are sinks of f_n, i.e., $d_{f_n}^+(x) = 0$ for $x \neq a$.
- $\Omega \ominus (f_n + h_n)$ is unhindered.

Aharoni et al. [3, Lem. 6.10] stated Lem. 8 with "unhindered" replaced by "loose". Their proof is structurally similar to ours, but assumes that $\Omega \ominus (f_n + h_n)$ is loose

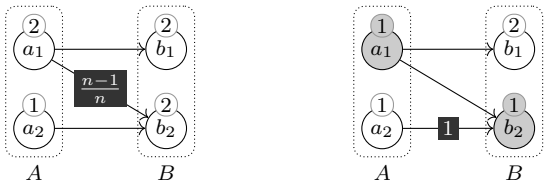

Fig. 16 A loose bipartite web Ω with a sequence of currents f_n (left) and the residual web $\Omega_\infty = \Omega \ominus (\lim_{n\to\infty} f_n)$ of the limit flow (right), which is not loose as shown by the non-zero wave.

rather than unhindered. Yet, taking the limit preserves only unhinderedness, not looseness. For example, Fig. 16 shows a loose bipartite web on the left. Suppose that we want to saturate the vertex a_1 and suppose that the saturation function *sat* always picks b_2 as the neighbour vertex whose weight should be reduced. Then, we can get a sequence of webs $(\Omega_n)_{n\in\mathbb{N}} = \Omega \ominus (f_n + h_n)$ with weight reductions on b_2 given by $w_{\Omega_n}(b_2) = 2 - \frac{n-1}{n}$, corresponding to the currents f_n shown in Fig. 16. Since Ω_n is loose, the waves h_n are always the zero wave. In the limit $n \to \infty$, the residual web $\Omega_\infty = \Omega \ominus (\lim_{n\to\infty}(f_n + h_n))$ is not loose as shown by the wave in Fig. 16 on the right. Our proof does not suffer from this problem because our induction invariant is unhinderedness rather than looseness.

7 Application: Relation Lifting over Discrete Probability Distributions

We now formalize an application of the max-flow min-cut theorem: the characterization of the relator for discrete probability distributions. We follow Sack and Zhang's proof [33], which generalizes the proofs by Desharnais [14] and Baier et al. [5] from finite to discrete distributions. This relator plays an important role in probabilistic bisimulation [13], probabilistic relational Hoare logic [10], and bounded natural functors [17]. At the end of this section, we highlight two use cases of the characterization theorem (Sects. 7.1 and 7.2).

Definition 16 (Probability mass function) A discrete probability distribution over elemenary events Ω is given by a probability mass function $X :: \Omega \to \mathbb{R}_{\geq 0}$ such that $\sum_{\omega\in\Omega} X(\omega) = 1$. So $X(x)$ denotes the probability mass for each elementary event x. An event $A \subseteq \Omega$ has then probability $\mathcal{P}[X \in A] = \sum_{x\in A} X(x)$. The support $\mathsf{supp}(X)$ of X is the set of elementary events with positive mass: $\mathsf{supp}(X) = \{x \in \Omega \mid X(x) > 0\}$. This set is countable.

Definition 17 (Marginal) For a discrete probability distribution over pairs of elementary events with probability mass function $X :: \Omega_1 \times \Omega_2 \to \mathbb{R}_{\geq 0}$, the marginals $\pi_1 X$ and $\pi_2 X$ are given by $\pi_1 X(x) = \mathcal{P}[X \in \{x\} \times \Omega_2]$ and $\pi_2 X(y) = \mathcal{P}[X \in \Omega_1 \times \{y\}]$.

The relator for discrete probability distributions transforms a relation R between elementary events into a relation $R^\uparrow$ between discrete probability distributions.

Definition 18 (Relator for probability distributions) Let $R \subseteq \Omega_1 \times \Omega_2$ be a relation between elementary events. The lifted relation $R^\uparrow$ relates two distributions

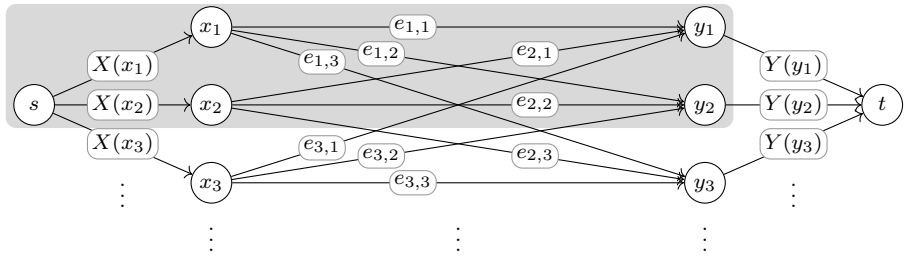

Fig. 17 Network for the characterization of the relator for discrete probability distributions and some cut (grey area).

X and Y over Ω_1 and Ω_2, respectively, iff X and Y are the marginals of some joint distribution Z over $\Omega_1 \times \Omega_2$ whose support is contained in R. Formally,

$$R^{\uparrow} = \{(X, Y) \mid \exists Z.\ X = \pi_1 Z \wedge Y = \pi_2 Z \wedge \mathsf{supp}(Z) \subseteq R\}.$$

We are interested in the following characterization theorem, where we let $R[A] = \{b \in \Omega_2 \mid \exists a \in A.\ a\ R\ b\}$ for a relation $R \subseteq \Omega_1 \times \Omega_2$ and $A \subseteq \Omega_1$.

Lemma 12 (Characterization for $\_^{\uparrow}$ [33, Lemma 1]) *Let $R \subseteq \Omega_1 \times \Omega_2$ and X and Y be two discrete probability distributions over Ω_1 and Ω_2, respectively. Then the following are equivalent:*

1. *$X\ R^{\uparrow}\ Y$*
2. *$\mathcal{P}[X \in A] \leq \mathcal{P}[Y \in R[A]]$ for all $A \subseteq \Omega_1$*

Proof The direction $1 \implies 2$ easily follows from the definitions. For the direction $2 \implies 1$, Sack and Zhang [33] consider the countable network $\Delta = (V, E, s, t, c)$ shown in Fig. 17. The vertices are the disjoint union of $\mathsf{supp}(X)$ and $\mathsf{supp}(Y)$ and additionally the source s and the sink t. Edges connect the source s to each $x_i \in \mathsf{supp}(X)$ and each $y_j \in \mathsf{supp}(Y)$ to the sink t; their capacities are given by the probability masses $X(x_i)$ and $Y(y_j)$ respectively. Moreover, there is an edge between every $x_i \in \mathsf{supp}(X)$ and $y_j \in \mathsf{supp}(Y)$. Its capacity $c(x_i, y_j) = e_{i,j}$ is defined as $e_{i,j} = 2$ if $x_i\ R\ y_j$ and 0 otherwise.[2]

By the max-flow min-cut theorem (Thm. 1), this network has an orthogonal flow f and cut C. The flow's value $|f|$ is at most 1 as $|f| = d_f^+(s) = \sum_{x \in \mathsf{supp}(X)} f(x) \leq \sum_{x \in \mathsf{supp}(X)} X(x) = 1$. Conversely, the cut C's value $|C|$ is at least 1 by the following analysis:

1. If $C = \{s\}$ or $\mathsf{supp}(Y) \subseteq C$, then $|C| \geq 1$ holds trivially, using $t \notin C$ in the latter case.
2. Otherwise consider $L = C \cap \mathsf{supp}(X)$. For the example cut in Fig. 17, we have $L = \{x_1, x_2\}$.
   (a) If $R[L] \not\subseteq C$, then some edge (x_i, y_j) leaves the cut C for some $x_i \in L$ and $y_j \notin C$ with $x_i\ R\ y_j$. (For example, (x_1, y_3) in Fig. 17 if we suppose $x_1\ R\ y_3$.) So $|C| \geq e_{i,j} = 2$.

[2] Sack and Zhang set $e_{i,j} = \infty$ if $x_i\ R\ y_j$, but the max-flow min-cut theorem handles only finite edge capacities. Their argument works unchanged for any value greater than 1, such as our choice of 2.

(b) If $R[L] \subseteq C$, then all edges (x_i, y_j) leaving C have capacity $e_{i,j} = 0$. (In Fig. 17, we would have neither $x_1\ R\ y_3$ nor $x_2\ R\ y_3$ and thus $e_{1,3} = e_{2,3} = 0$.) Then, $|C| = \left(\sum_{x \in \mathsf{supp}(X) - L} X(x)\right) + \left(\sum_{y \in R[L]} Y(y)\right) = \mathcal{P}[X \notin L] + \mathcal{P}[Y \in R[L]] \geq \mathcal{P}[X \notin L] + \mathcal{P}[X \in L] = 1$, where the inequality comes from the premise of the backwards direction.

Since Δ does not contain infinite paths, we get $|f| = |C|$ by flow preservation and thus $|f| = 1$. Set $Z(x_i, y_j) = f(x_i, y_j)$ for $x_i \in \mathsf{supp}(X)$ and $y_j \in \mathsf{supp}(Y)$ and $Z(x, y) = 0$ otherwise. Then Z is a probability mass function over $\Omega_1 \times \Omega_2$ and a witness to the existential in Def. 18. □

As is, this proof needs the unbounded max-flow min-cut theorem. Since R may relate an elementary event x_i to infinitely many neighbours $y_j \in R[\{x_i\}]$, the network Δ violates the condition of the bounded case (Sect. 4.1): $d_c^+(x_i) = \sum_{y_j \in R[\{x_i\}]} e_{i,j} = \sum_{y_j \in R[\{x_i\}]} 2 = \infty \not< \infty$.

Yet, we can tweak the capacity definition so that the modified network satisfies the condition of the bounded case. For $x_i\ R\ y_j$, we choose $e_{i,j} = Y(y_j)$ instead of $e_{i,j} = 2$. Then $d_c^+(x_i) = \sum_{y_j \in R[\{x_i\}]} e_{i,j} = \sum_{y_j \in R[\{x_i\}]} Y(y_j) = \mathcal{P}[Y \in R[\{x_i\}]] \leq 1$. The proof remains the same except for the argument that $|C| \geq 1$. The above case distinctions are all subsumed by the following general analysis. Every edge in the network originates either in some x_i or in some y_j or in s. We consider the following sets:

- $L = C \cap \mathsf{supp}(X)$ the origins x_i of edges (x_i, y_j) that might leave C.
- $K = C \cap \mathsf{supp}(Y)$ the origins y_j of edges (y_j, t) that leave C (as $t \notin C$).
- $S = \mathsf{supp}(X) - C$ the targets x_i of edges (s, x_i) that leave C (as $s \in C$).

(In Fig. 17, we have $L = \{x_1, x_2\}$ and $K = \{y_1, y_2\}$ and $S = \{x_3\}$.) Then

$$|C| = \underbrace{\sum_{x_i \in L} \sum_{y_j \in R[\{x_i\}] - C} c(x_i, y_j)}_{\geq \sum_{y_j \in R[L] - C} Y(y_j)} + \underbrace{\sum_{y_j \in K} c(y_j, t)}_{\geq \sum_{y_j \in C \cap R[L]} Y(y_j)} + \underbrace{\sum_{x_i \in S} c(s, x_i)}_{= \sum_{x_i \in S} X(x_i)}$$

For the inequality of the first sum, rather than adding $c(x_i, y_j) = Y(y_j)$ once for each of y_j's neighbours x_i in L, we add it only once for each y_j. In the second sum, $c(y_j, t) = Y(y_j)$ and $K \supseteq C \cap R[L]$ justifies the inequality. In the third sum, we use $c(s, x_i) = X(x_i)$. So

$$\begin{aligned} |C| &\geq \mathcal{P}[Y \in R[L] - C] + \mathcal{P}[Y \in R[L] \cap C] + \mathcal{P}[X \in S] \\ &= \mathcal{P}[Y \in R[L]] + \mathcal{P}[X \notin L] \\ &\geq \mathcal{P}[X \in L] + \mathcal{P}[X \notin L] = 1. \end{aligned}$$

We have formalized both Sack and Zhang's proof and the modified proof of Lemma 12 for the relator rel_pmf from Isabelle/HOL's probability theory library.[3] Interestingly, the analysis of the modified network requires the about same number of proof lines as the proof with $c_{i,j} = 2$.

[3] Sack and Zhang's proof is formalized in the theory `Rel_PMF_Characterisation` in the accompanying AFP entry [25] version for Isabelle2016-1. The modified proof can be found in the current development version at https://devel.isa-afp.org/browser_info/current/AFP/MFMC_Countable/Rel_PMF_Characterisation_MFMC.html.

7.1 Distributivity over relation composition

Let $R \bullet S = \{(x, z) \mid \exists y.\ x\ R\ y \wedge y\ S\ z\}$ denote the composition of the relations R and S. The relator $\_^{\uparrow}$ distributes over relation composition, i.e., $(R \bullet S)^{\uparrow} = R^{\uparrow} \bullet S^{\uparrow}$. So $\_^{\uparrow}$ satisfies the functor laws, interpreted in the category of relations, as the other functor law $(=)^{\uparrow} = (=)$ holds trivially. Distributivity over relation composition is also required for discrete probabilities being a bounded natural functor (BNF) in the category of sets. BNFs are the modular building blocks for constructing datatypes in Isabelle/HOL [11].

Typically, the difficult part is the direction from right to left: If $X\ R^{\uparrow}\ Y$ and $Y\ S^{\uparrow}\ Z$, then $X\ (R \bullet S)^{\uparrow}\ Z$. In previous work [17], we have formally derived this directly from the definition (Def. 18), where we were able to cut the proof from initially 577 lines down to 46 lines with substantial effort. In contrast, Lemma 12 makes this proof trivial: for all A, we have $\mathcal{P}[X \in A] \leq \mathcal{P}[Y \in R[A]] \leq \mathcal{P}[Z \in S[R[A]]]$ and $S[R[A]] = (R \bullet S)[A]$.

7.2 Parametricity of fixpoints of subprobabilities

Unlike a probability distribution, a subprobability distribution X may leave some probability mass unassigned, i.e., $\sum_{x \in \Omega} X(x) \leq 1$. In Isabelle/HOL's probability library, a discrete subprobability distribution X over Ω is represented by a probability mass function $X :: \Omega^{\perp} \to \mathbb{R}_{\geq 0}$ where $\_^{\perp}$ adds a new element $\perp$ to a set; this element $\perp$ absorbs the probability mass that is not assigned to Ω. We write $\mathbb{D}(\Omega)$ for the set of all subprobability distributions over Ω. The corresponding relator $\_^{\Uparrow}$ lifts a relation R between Ω_1 and Ω_2 to subprobability distributions over Ω_1 and Ω_2; it is given by $R^{\Uparrow} = \{(X, Y) \mid X\ (R^{\perp})^{\uparrow}\ Y\}$ where $R^{\perp}$ extends R with the pair $(\perp, \perp)$.

Lemma 12 yields the following characterization of $\_^{\Uparrow}$ as a corollary:

Corollary 1 *For a relation $R \subseteq \Omega_1 \times \Omega_2$ and subprobability distributions $X :: \mathbb{D}(\Omega_1)$ and $Y :: \mathbb{D}(\Omega_2)$, the following are equivalent:*

1. $X\ R^{\Uparrow}\ Y$
2. $\mathcal{P}[Y \in \Omega_2] \leq \mathcal{P}[X \in \Omega_1]$ *and* $\mathcal{P}[X \in A] \leq \mathcal{P}[Y \in R[A]]$ *for all* $A \subseteq \Omega_1$.

Subprobability distributions are partially ordered by $X \sqsubseteq Y$ iff $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$, i.e., every elementary event ω other than the artificial $\perp$ has the same or higher probability with Y than X. This ordering is chain-complete, i.e., every chain $\mathcal{X}$ has a least upper bound $\mathsf{sup}\,\mathcal{X}$. The least element is the subprobability distribution $\mathbf{0}$ that assigns no probability mass to Ω. Measuring the probability of an event is chain-continuous: $\mathcal{P}[\mathsf{sup}\,\mathcal{X} \in A] = \mathsf{sup}_{X \in \mathcal{X}}\, \mathcal{P}[X \in A]$ for a chain $\mathcal{X}$ of subprobability distributions. Chain-completeness yields a least fixpoint operator $\mathsf{fix} :: (\mathbb{D}(\Omega) \xrightarrow{\mathsf{m}} \mathbb{D}(\Omega)) \to \mathbb{D}(\Omega)$ via transfinite fixpoint iteration starting from $\mathbf{0}$ (and similarly for functions that return subprobability distributions), where $\xrightarrow{\mathsf{m}}$ denotes the space of monotone functions.

The next proposition states that fix preserves relation lifting. This preservation property forms the cornerstone for proving that recursively defined subprobabilities are relationally parametric, which itself is used for example by Isabelle's Transfer [18] and Types-To-Sets packages [21]. In [27], we define a probabilistic loop operator as a fixpoint and prove it being parametric this way.

Proposition 3 *The fixpoint operator* fix *preserves relation lifting. Let* $f :: \mathbb{D}(\Omega_1) \xrightarrow{m} \mathbb{D}(\Omega_1)$ *and* $g :: \mathbb{D}(\Omega_2) \xrightarrow{m} \mathbb{D}(\Omega_2)$ *such that* $X\ R^{\Uparrow}\ Y$ *implies* $f(X)\ R^{\Uparrow}\ g(Y)$ *for all* X *and* Y*. Then* $\mathsf{fix}(f)\ R^{\Uparrow}\ \mathsf{fix}(g)$*.*

Proof By parallel induction on the two fixpoints. The base case $\mathbf{0}\ R^{\Uparrow}\ \mathbf{0}$ is trivial and the inductive step is exactly by assumption. The interesting part is the transfinite inductive step: If $\mathcal{X} = X_0, X_1, X_2, \ldots$ and $\mathcal{Y} = Y_1, Y_2, Y_3, \ldots$ are chains of subprobability distributions over Ω_1 and Ω_2 such that $X_i\ R^{\Uparrow}\ Y_i$ for all i, we must show that $(\mathsf{sup}\,\mathcal{X})\ R^{\Uparrow}\ (\mathsf{sup}\,\mathcal{Y})$, too. With Cor. 1, this directly follows from order-continuity of measuring the probability of an event:

$$\mathcal{P}[\mathsf{sup}\,\mathcal{Y} \in \Omega_2] = \mathsf{sup}_i\,\mathcal{P}[Y_i \in \Omega_2] \leq \mathsf{sup}_i\,\mathcal{P}[X_i \in \Omega_1] = \mathcal{P}[\mathsf{sup}\,\mathcal{X} \in \Omega_1]$$
$$\mathcal{P}[\mathsf{sup}\,\mathcal{X} \in A] = \mathsf{sup}_i\,\mathcal{P}[X_i \in A] \leq \mathsf{sup}_i\,\mathcal{P}[Y_i \in R[A]] = \mathcal{P}[\mathsf{sup}\,\mathcal{Y} \in R[A]] \qquad \square$$

Our formalization of this proof lives in the theory `SPMF.thy` in Isabelle/HOL's probability library, with the $2 \implies 1$ direction of Lemma 12 as an additional assumption using an Isabelle locale (changeset 1bc6816fd525, lines 1697–1823). We discharge this assumption in our AFP entry [25].

8 Related work

Lee [24] and Lammich and Sefidgar [22, 23] have formalized the max-flow min-cut theorem for *finite* networks in Mizar and Isabelle/HOL, respectively. Lammich and Sefidgar additionally formalize and verify several max-flow algorithms. We reused Lammich and Sefidgar's formalization in our proof of Prop. 1. We make no algorithmic considerations, as countable networks are infinite objects that lie beyond the reach of traditional notions of algorithms.

Lyons and Peres [30, Thm. 3.1] consider countable locally finite networks, where every vertex has only finitely many neighbours, and without a sink. They show that the maximum flow's value equals the value of a minimum cut, where a cut here contains an edge of every infinite simple path that starts at the source. Like our proof for the bounded case, their proof extends the max-flow min-cut theorem for finite networks using majorised convergence. Since their graphs are locally finite, all summations of interest are finite by construction.

We have already mentioned related work on the characterization of the relator for (sub)probability distributions in Sect. 7. The related work on the distributivity proof for $\_^{\uparrow}$ is discussed in detail in [17].

CertiCrypt [10] formalizes a probabilistic while loop as a fixpoint using the Coq library ALEA [4]. Deriving the pRHL while rule is equivalent to showing that while is parametric. Barthe et al.'s proof manually constructs the witness for the existential in Def. 18 explicitly for while, by defining a coupled execution of the related while programs. Our fixpoint rule works for general fixpoints. In fact, we have derived parametricity for a probabilistic while loop from Prop. 3 by just unfolding definitions [27].

Barthe et al. [9] have generalized the relator $\_^{\uparrow}$ to an approximate lifting operation called $\star$-lifting. They derive a corresponding characterization using an analogous argument. Like Sack and Zhang, they wrongly set $e_{i,j} = \infty$ instead of using a sufficiently large finite value. Moreover, we can tweak the edge capacities in their network as we did for Sack and Zhang's in Sect. 7 so that every vertex's outgoing total capacity is finite. Formalizing this generalization is left as future work.

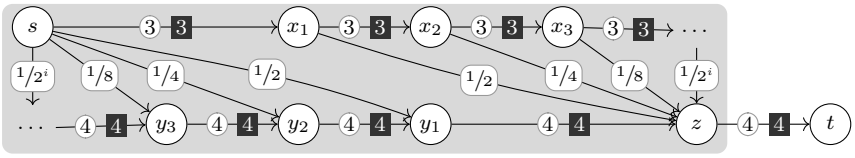

Fig. 18 An infinite network with an orthogonal pair of a cut and a flow.

9 Conclusion

In this paper, we have formalized a strong max-flow min-cut theorem for countable networks in Isabelle/HOL. To rule out anomalities due to the network being infinite, the theorem statement avoids imprecise infinite sums and instead compares the saturation edge by edge. During the formalization, we have discovered and fixed a number of problems in the original proof [3].

Arguably, this statement still does not capture the intuition fully. For example, the infinite network in Fig. 18 has a cut of value 4 with an orthogonal flow. This is the cut that the proof of Thm. 1 constructs. Yet, this cut is not minimal: The cut that separates the upper nodes from the lower nodes would be saturated by a flow of 2 units (not shown). This illustrates the intricacies of infinite networks: The out-flow from the source s of value 3 drains away in the infinite ray $s \to x_1 \to x_2 \to x_3 \to \ldots$. Conversely, the in-flow to the sink t of value 4 is pulled in via the infinite path $\ldots \to y_3 \to y_2 \to y_1 \to z \to t$. So this network shows that the outflow from the source may exceed the capacity of a cut and yet not saturate it.

Aharoni et al. [3, Sects. 7–8] study two restrictions on networks that avoid such anomalies: networks without infinite edge-disjoint paths and locally-finite networks. We have not yet formalized these results. Neither result applies to the network in Fig. 18. So finding a more intuitive statement of the max-flow min-cut theorem for countable networks is still an open problem.

Acknowledgements Swiss National Science Foundation grant 153217 "Formalising Computational Soundness for Protocol Implementations". This work was partially done while the author was at ETH Zurich.

We thank Ron Aharoni and Eli Berger for helping to clarify the weaknesses in the original proofs. S. Reza Sefidgar and the anonymous reviewers helped to improve the presentation.

References

1. Aharoni, R.: Menger's theorem for graphs containing no infinite paths. European Journal of Combinatorics **4**, 201–204 (1983). DOI 10.1016/S0195-6698(83)80012-2
2. Aharoni, R., Berger, E.: Menger's theorem for infinite graphs. Inventiones mathematicae **176**(1), 1–62 (2009). DOI 10.1007/s00222-008-0157-3
3. Aharoni, R., Berger, E., Georgakopoulos, A., Perlstein, A., Sprüssel, P.: The max-flow min-cut theorem for countable networks. Journal of Combinatorial Theory, Series B **101**, 1–17 (2010). DOI 10.1016/j.jctb.2010.08.002
4. Audebaud, P., Paulin-Mohring, C.: Proofs of randomized algorithms in Coq. Science of Computer Programming **74**, 568—589 (2009). DOI 10.1016/j.scico.2007.09.002
5. Baier, C., Engelen, B., Majster-Cederbaum, M.: Deciding bisimilarity and similarity for probabilistic processes. Journal of Computer and System Sciences **60**(1), 187–231 (2000). DOI 10.1006/jcss.1999.1683

6. Ballarin, C.: Locales and locale expressions in Isabelle/Isar. In: S. Berardi, M. Coppo, F. Damiani (eds.) TYPES 2003, *LNCS*, vol. 3085, pp. 34–50. Springer (2004). DOI 10.1007/978-3-540-24849-1_3
7. Ballarin, C.: Locales: A module system for mathematical theories. Journal of Automated Reasoning **52**, 123–153 (2014). DOI 10.1007/s10817-013-9284-7
8. Ballarin, C.: Exploring the structure of an algebra text with locales. Journal of Automated Reasoning **64**, 1093–1121 (2020). DOI 10.1007/s10817-019-09537-9
9. Barthe, G., Espitau, T., Hsu, J., Sato, T., Strub, P.Y.: *-liftings for differential privacy. In: I. Chatzigiannakis, P. Indyk, F. Kuhn, A. Muscholl (eds.) ICALP 2017, *LIPIcs*, vol. 80, pp. 102:1–102:12. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2017). DOI 10.4230/LIPIcs.ICALP.2017.102
10. Barthe, G., Grégoire, B., Zanella Béguelin, S.: Formal certification of code-based cryptographic proofs. In: POPL 2009, pp. 90–101. ACM (2009). DOI 10.1145/1480881.1480894
11. Blanchette, J.C., Hölzl, J., Lochbihler, A., Panny, L., Popescu, A., Traytel, D.: Truly modular (co)datatypes for Isabelle/HOL. In: ITP 2014, *LNCS*, vol. 8558, pp. 93–110. Springer (2014). DOI 10.1007/978-3-319-08970-6_7
12. Bourbaki, N.: Sur le théorème de Zorn. Archiv der Mathematik **2**(6), 434–437 (1949)
13. Deng, Y.: Semantics of Probabilistic Processes. Springer (2014). DOI 10.1007/978-3-662-45198-4
14. Desharnais, J.: Labelled Markov processes. Ph.D. thesis, McGill University (1999)
15. Edmonds, J., Karp, R.M.: Theoretical improvements in algorithmic efficiency for network flow problems. Journal of the ACM **19**(2), 248–264 (1972). DOI 10.1145/321694.321699
16. Ford, L.R., Fulkerson, D.R.: Maximal flow through a network. Canadian Journal of Mathematics **8**, 399–404 (1956). DOI 10.4153/CJM-1956-045-5
17. Hölzl, J., Lochbihler, A., Traytel, D.: A formalized hierarchy of probabilistic system types. In: C. Urban, X. Zhang (eds.) ITP 2015, *LNCS*, vol. 9236, pp. 203–220. Springer (2015). DOI 10.1007/978-3-319-22102-1_13
18. Huffman, B., Kunčar, O.: Lifting and Transfer: A modular design for quotients in Isabelle/HOL. In: CPP 2013, *LNCS*, vol. 8307, pp. 131–146. Springer (2013). DOI 10.1007/978-3-319-03545-1_9
19. Immler, F.: Generic construction of probability spaces for paths of stochastic processes in Isabelle/HOL. Master's thesis, Fakultät für Informatik, Technische Universität München (2012)
20. Kellerer, H.G.: Funktionen auf Produkträumen mit vorgegebenen Marginal-Funktionen. Mathematische Annalen **144**, 323–344 (1961). DOI 10.1007/BF01470505
21. Kunčar, O., Popescu, A.: From types to sets by local type definition in higher-order logic. Journal of Automated Reasoning **62**, 237—260 (2019). DOI 10.1007/s10817-018-9464-6
22. Lammich, P., Sefidgar, S.R.: Formalizing the Edmonds-Karp algorithm. In: J.C. Blanchette, S. Merz (eds.) ITP 2016, *LNCS*, vol. 9807, pp. 219–234. Springer (2016). DOI 10.1007/978-3-319-43144-4_14
23. Lammich, P., Sefidgar, S.R.: Formalizing network flow algorithms: A refinement approach in Isabelle/HOL. Journal of Automated Reasoning **62**, 261–280 (2019). DOI 10.1007/s10817-017-9442-4
24. Lee, G.: Correctnesss of Ford-Fulkerson's maximum flow algorithm. Formalized Mathematics **13**(2), 305–314 (2005). URL https://fm.mizar.org/2005-13/pdf13-2/glib_005.pdf
25. Lochbihler, A.: A formal proof of the max-flow min-cut theorem for countable networks. Archive of Formal Proofs (2016). http://www.isa-afp.org/entries/MFMC_Countable.shtml, Formal proof development
26. Lochbihler, A.: Probabilistic functions and cryptographic oracles in higher-order logic. In: P. Thiemann (ed.) ESOP 2016, *LNCS*, vol. 9632, pp. 503–531. Springer (2016). DOI 10.1007/978-3-662-49498-1_20
27. Lochbihler, A.: Probabilistic while loop. Archive of Formal Proofs (2017). https://isa-afp.org/entries/Probabilistic_While.html, Formal proof development
28. Lochbihler, A.: A mechanized proof of the max-flow min-cut theorem for countable networks. In: L. Cohen, C. Kaliszyk (eds.) ITP 2021, *LIPIcs*, vol. 193, pp. 25:1–25:18 (2021). DOI 10.4230/LIPIcs.ITP.2021.25
29. Lochbihler, A.: A mechanized proof of the max-flow min-cut theorem for countable networks with applications to probability theory. http://www.andreas-lochbihler.de/pub/lochbihler-mfmc.pdf (2021)
30. Lyons, R., Peres, Y.: Probability on Trees and Networks. Cambridge University Press, New York (2017). DOI 10.1017/9781316672815

31. Naraschewski, W., Wenzel, M.: Object-oriented verification based on record subtyping in higher-order logic. In: J. Grundy, M. Newey (eds.) TPHOLs 1998, *LNCS*, vol. 1479, pp. 349–366. Springer (1998). DOI 10.1007/BFb0055146
32. Sabot, C., Tournier, L.: Random walks in Dirichlet environment: an overview. Annales de la Faculté des sciences de Toulouse: Mathématiques **Ser. 6, 26**(2), 463–509 (2017). DOI 10.5802/afst.1542
33. Sack, J., Zhang, L.: A general framework for probabilistic characterizing formulae. In: V. Kuncak, A. Rybalchenko (eds.) VMCAI 2012, *LNCS*, vol. 7148, pp. 396–411. Springer (2012). DOI 10.1007/978-3-642-27940-9_26
34. Smolka, G., Schäfer, S., Doczkal, C.: Transfinite constructions in classical type theory. In: C. Urban, X. Zhang (eds.) ITP 2015, *LNCS*, vol. 9236, pp. 391–404. Springer (2015). DOI 10.1007/978-3-319-22102-1_26
35. Strassen, V.: The existence of probability measures with given marginals. Annals of Mathematical Statistics **36**(2), 423–439 (1965). DOI 10.1214/aoms/1177700153
36. Wiedijk, F.: The de Bruijn factor. https://www.cs.ru.nl/~freek/factor/factor.pdf (2000)