# Supplementary Document

Bhat et al. *Billi: Provably Accurate and Scalable Bubble Detection in Pangenome Graphs*

## Contents



---

# A Properties of panbubbles and hairpins: full exposition

## Proof of Lemma 1

**Lemma A.1.** *A vertex can be the entrance vertex for at most one panbubble.*

*Proof.* Proof by contradiction. Suppose there exist two distinct panbubbles $\langle s, t_1 \rangle$, $\langle s, t_2 \rangle$, $t_1 \neq t_2$.

If $t_2 \in U(s, t_1)$, then $t_2$ is also in $B(s, t_1)$. However, this contradicts the minimality condition for $\langle s, t_1 \rangle$. Similarly, $t_1$ being a vertex in $U(s, t_2)$ also results in a contradiction.

Now suppose that $t_2 \notin U(s, t_1)$. Since $t_2 \in U(s, t_2)$, $t_2$ is reachable from $s$ using some walk that does not pass through $s$ or $t_2$ (from the definition of $U(s, t_2)$). This means there exists a walk $\omega$ between $s$ and $t_2$ in which both $s$ and $t_2$ appear only once. Because $t_2$ appears in $\omega$, it is correct to state that not all vertices in walk $\omega$ belong to $U(s, t_1)$. Consider the ordered list of the vertices in walk $\omega$ starting from vertex $s$. Observe that the vertex just before the first vertex in $\omega$ that is not in $U(s, t_1)$ must be either $s$ or $t_1$. But $s$ cannot appear twice in $\omega$, so it must be $t_1$. But the appearance of $t_1$ in $\omega$ implies that $t_1$ is in $U(s, t_2)$, which violates the first half of the argument. □

**Lemma 1.** *In a compact biedged graph $G_b = (V_b, E_b, f_b)$, the number of distinct panbubbles can be at most $|V_b/2|$. The number of distinct hairpins can be at most $|V_b|$.*

*Proof.* A vertex cannot be the entrance vertex for two different panbubbles (Lemma A.1), so each entrance is unique. In a compact biedged graph, there are $|V_b|$ vertices, but each panbubble is defined by two entrance vertices, so the number of panbubbles can be at most $|V_b/2|$.

Since hairpins are defined by a single entrance vertex, each vertex can correspond to at most one hairpin, and therefore the number of distinct hairpins can be at most $|V_b|$. □

## Proof of Lemma 2

Recall from the computation of the auxiliary undirected multigraph $G_U$ and its depth-first spanning tree $T_{v_{src}}$ that every edge of $G_U$ must be either a tree edge or a back edge. Furthermore, every edge of $G_U$ derived from a black edge of $G_b$ must be a tree edge. The following two corollaries are a direct consequence of this.

**Corollary A.1.** *For all $\{v_1, v_2\} \in E_U$ such that $v_1 \neq v_2$, either $v_1 \prec v_2$ or $v_2 \prec v_1$.*

**Corollary A.2.** *For all $v \in V_b$, there exists a root-leaf path in $T_{v_{src}}$ that contains $\bar{v}$, $\overline{\hat{v}}$ and edge $\bar{e}_v = \{\bar{v}, \overline{\hat{v}}\}$.*

For a panbubble $\langle s, t\rangle$ in $G_b$, let $\overline{U}(s,t)$ denote the set of vertices in $G_U$ corresponding to the set $U(s,t)$ in $G_b$. Formally, $\overline{U}(s,t) = \{\bar{v} | v \in U(s,t)\}$. Similarly, we define $\overline{B}(s,t) = \{\bar{v} | v \in B(s,t)\}$.

**Lemma A.2.** *For each panbubble $\langle s, t\rangle$ in $G_b$, the set of vertices in $G_U$ reachable from $\overline{\hat{s}}$ without passing through $\bar{s}$ or $\bar{t}$ equals $\overline{U}(s,t)$.*

*Proof.* Let $X$ denote the set of vertices reachable from $\overline{\hat{s}}$ without passing through $\bar{s}$ or $\bar{t}$. We need to show that $X = \overline{U}(s,t)$. We first argue that $\overline{U}(s,t) \subseteq X$. Recall that for every $v \in U(s,t)$, $v$ is reachable from $s$ without passing through $s$ or $t$ (from the definition of $U(s,t)$). The corresponding walk in $G_U$ contains a subwalk from $\overline{\hat{s}}$ to $\bar{v}$ that does not pass through $\bar{s}$ or $\bar{t}$. Therefore, $\bar{v} \in X$.

Next, we prove $X \subseteq \overline{U}(s,t)$ by contradiction. Suppose that $X \not\subseteq \overline{U}(s,t)$. This implies the existence of a vertex in $G_U$ such that (i) this vertex is reachable from $\overline{\hat{s}}$ without passing through $\bar{s}$ or $\bar{t}$, and (ii) this vertex is not in $\overline{U}(s,t)$. In that case, there must exist at least one walk $\omega$ having vertex sequence $(\overline{\hat{s}}, v_1, v_2, \ldots, v_k)$ in $G_U$ with $k \geq 1$ such that $\bar{s}$ and $\bar{t}$ do not appear in $\omega$, and $v_k$ is the only vertex in $\omega$ that does not belong to $\overline{U}(s,t)$. It follows that vertices $v_1, v_2, \ldots, v_{k-1}$ are in $\overline{B}(s,t)$. Further note that $v_k \neq v_{src}$, because if $v_k = v_{src}$, then we would have a tip vertex in $B(s,t)$, which contradicts the contiguity condition of $\langle s, t\rangle$.

Consider the case when $v_k \neq v_{src}$ and $k > 1$. Since $v_{k-1} \in \overline{B}(s,t)$ and $v_k \neq v_{src}$, we must have vertices corresponding to $v_{k-1}$ and $v_k$ in $G_b$. We must also have an edge in $G_b$ between $v_{k-1}$ and $v_k$. Since $v_{k-1} \in \overline{B}(s,t)$ and $v_k \notin \overline{U}(s,t)$, observe that the presence of an edge in $G_b$ between a vertex in $B(s,t)$ and a vertex in $V_b \setminus U(s,t)$ contradicts the separable condition of $\langle s, t\rangle$.

The other case, i.e., $v_k \neq v_{src}$ and $k = 1$, implies the presence of an edge in $G_b$ between $\hat{s}$ and a vertex in $V_b \setminus U(s,t)$, again contradicting the separable condition of $\langle s, t\rangle$. ☐

**Corollary A.3.** *For each panbubble $\langle s, t\rangle$ in $G_b$, the set of vertices in $G_U$ reachable from $\overline{\hat{s}}$ without passing through $\bar{s}$ or $\bar{t}$ does not contain $v_{src}$.*

Next, we analyze the properties of edges $\bar{e}_s$ and $\bar{e}_t$ in $G_U$.

**Lemma A.3.** *For each panbubble $\langle s, t\rangle$ in $G_b$, either $\bar{e}_s \prec \bar{e}_t$ or $\bar{e}_t \prec \bar{e}_s$ in $G_U$.*

*Proof.* Proof by contradiction. Suppose neither $\bar{e}_s \prec \bar{e}_t$ nor $\bar{e}_t \prec \bar{e}_s$. Then there does not exist any root-leaf path in $T_{v_{src}}$ that contains both $\bar{e}_s$ and $\bar{e}_t$. Suppose wlog that $\bar{s}$ was marked as visited before $\bar{t}$ during the depth-first traversal from vertex $v_{src}$. By Corollary A.1 and A.2, we have either $\bar{s} \prec \overline{\hat{s}}$ or $\overline{\hat{s}} \prec \bar{s}$. We consider these two cases one by one.

Case 1 ($\bar{s} \prec \overline{\hat{s}}$): In our depth-first traversal, $\overline{\hat{s}}$ must have been visited immediately after $\bar{s}$. Since $t \in U(s,t)$, there exists a walk $(\bar{s}, \overline{\hat{s}}, \ldots, \bar{t})$ in $G_U$ not passing through $\bar{s}$ or $\bar{t}$. This means vertex $\bar{t}$ (and thus edge $\bar{e}_t$) must lie within the sub-tree of $\overline{\hat{s}}$ in $T_{v_{src}}$ This contradicts our initial assumption that there is no root-leaf path in $T_{v_{src}}$ that contains both $\bar{e}_s$ and $\bar{e}_t$.

Case 2 ($\overline{\hat{s}} \prec \bar{s}$): In the depth-first traversal, $\bar{s}$ must have been visited immediately after $\overline{\hat{s}}$ . Since $\bar{t}$ is marked as visited after $\bar{s}$ during the traversal, $\bar{t}$ is certainly not an ancestor of $\bar{e}_s$. As a result, there exists a path from $\overline{\hat{s}}$ to $v_{src}$ in $G_U$ without passing through $\bar{s}$ or $\bar{t}$. This contradicts our previous claim in Corollary A.3. ☐

**Lemma A.4.** *For each panbubble $\langle s, t\rangle$ in $G_b$, $\bar{e}_s \prec \bar{e}_t$ if and only if $\bar{s} \prec \overline{\hat{s}} \prec \overline{\hat{t}} \prec \bar{t}$.*

*Proof.* [$\Leftarrow$] $\bar{s} \prec \overline{\hat{s}} \prec \overline{\hat{t}} \prec \bar{t}$ directly implies $\bar{e}_s \prec \bar{e}_t$.

[$\Rightarrow$] If $\bar{e}_s \prec \bar{e}_t$, then clearly $\bar{s} \prec \bar{t}$, $\bar{s} \prec \overline{\hat{t}}$, $\overline{\hat{s}} \prec \bar{t}$, and $\overline{\hat{s}} \prec \overline{\hat{t}}$. Next we will argue that $\bar{s} \prec \overline{\hat{s}}$. Assume for contradiction that $\overline{\hat{s}} \prec \bar{s}$. In that case, there would exist a path from $\overline{\hat{s}}$ to $v_{src}$ without passing through $\bar{s}$ or $\bar{t}$, which contradicts our earlier claim in Corollary A.3. Next we prove $\overline{\hat{t}} \prec \bar{t}$ by contradiction. Suppose $\bar{t} \prec \overline{\hat{t}}$. Then $\bar{s} \prec \overline{\hat{s}} \prec \bar{t} \prec \overline{\hat{t}}$ and there exists a path $p = (\bar{s}, \overline{\hat{s}}, \ldots, \bar{t})$ in $G_U$ which does not include $\overline{\hat{t}}$. Recall that the vertices in a path are distinct by definition. Using Lemma A.2, every vertex on path $p$ belongs to $\overline{U}(s,t)$. Tracing this path back to $G_b$, vertex $t$ must connect to some vertex in $B(s,t)$ aside from $\hat{t}$. This contradicts the separable condition of panbubble $\langle s,t \rangle$. ☐

Using a symmetric argument, we obtain the following:

**Corollary A.4.** *For each panbubble $\langle s,t \rangle$ in $G_b$, $\bar{e}_t \prec \bar{e}_s$ if and only if $\bar{t} \prec \bar{\hat{t}} \prec \bar{\hat{s}} \prec \bar{s}$.*

These lemmas are all that are required to prove Lemma 2.

**Lemma 2.** *For every panbubble $\langle s,t \rangle$ in $G_b$, either $\bar{s} \prec \bar{\hat{s}} \prec \bar{\hat{t}} \prec \bar{t}$ or $\bar{t} \prec \bar{\hat{t}} \prec \bar{\hat{s}} \prec \bar{s}$ in $G_U$.*

*Proof.* For a panbubble $\langle s,t \rangle$ in $G_b$, we know either $\bar{e}_s \prec \bar{e}_t$ or $\bar{e}_t \prec \bar{e}_s$ in $G_U$ (Lemma A.3). From Lemma A.4, and Corollary A.4, we have that either $\bar{s} \prec \bar{\hat{s}} \prec \bar{\hat{t}} \prec \bar{t}$ or $\bar{t} \prec \bar{\hat{t}} \prec \bar{\hat{s}} \prec \bar{s}$ in $G_U$. ☐

## Proof of Lemma 3

**Lemma A.5.** *Tree edges $e_1$ and $e_2$ are cycle equivalent in $G_U$ if and only if they have the same set of brackets.*

*Proof.* This property holds for all undirected graphs; see Theorem 5 in [13]. ☐

**Lemma A.6.** *For each panbubble $\langle s,t \rangle$ in $G_b$, edges $\bar{e}_s$ and $\bar{e}_t$ have the same bracket set in $G_U$.*

*Proof.* Proof by contradiction. Assume that edges $\bar{e}_s$ and $\bar{e}_t$ have different bracket sets. Using Lemma A.3, either $\bar{e}_s \prec \bar{e}_t$ or $\bar{e}_t \prec \bar{e}_s$. Suppose wlog that $\bar{e}_s \prec \bar{e}_t$. Accordingly, $\bar{s} \prec \bar{\hat{s}} \prec \bar{\hat{t}} \prec \bar{t}$ (using Lemma A.4). Let $D_s$ and $D_t$ denote the set of all descendant vertices of edges $\bar{e}_s$ and $\bar{e}_t$, respectively. Let $A_s$ and $A_t$ denote the set of all ancestor vertices of the edges $\bar{e}_s$ and $\bar{e}_t$, respectively. For $\bar{e}_s$ and $\bar{e}_t$ to have an unequal bracket set, an edge must exist in $G_U$ such that either (a) one endpoint of the edge is in $D_s \setminus D_t$ and the other endpoint is in $A_s$, or (b) one end-point of the edge is in $D_t$ and the other endpoint is in $A_t \setminus A_s$.

Suppose the edge satisfies Condition (a). By Corollary A.3, the endpoint of this edge in $A_s$ cannot be $v_{src}$. Now, observe that using this edge, we have a path between $\bar{\hat{s}}$ and $\bar{s}$ in $G_U$ without using $\bar{e}_s$ or $\bar{e}_t$. Mapping this path back to the corresponding sequence of edges in $G_b$ contradicts the separable condition of $\langle s,t \rangle$. Similarly, if the edge satisfies Condition (b), then there exists a path between $\bar{\hat{s}}$ and $\bar{t}$ in $G_U$ without using $\bar{e}_s$ or $\bar{e}_t$. Considering the corresponding sequence of edges in $G_b$ again contradicts the separable condition of $\langle s,t \rangle$. ☐

From these two lemmas, we can prove the following:

**Lemma 3.** *For every panbubble $\langle s,t \rangle$ in $G_b$, edges $\bar{e}_s$ and $\bar{e}_t$ are cycle-equivalent in $G_U$.*

*Proof.* For a panbubble $\langle s,t \rangle$ in $G_b$, $\bar{e}_s$ and $\bar{e}_t$ have the same bracket set in $G_U$ (Lemma A.6) and hence are also cycle equivalent in $G_U$ (Lemma A.5). ☐

## Proof of Lemma 4

The following lemma can be proved by using the properties of a panbubble.

**Lemma 4.** *Suppose $\langle s,t\rangle$ is a panbubble in $G_b$. Consider walks in $G_b$ that do not pass through $s$ or $t$. For every vertex $u \in U(s,t)$, there exist such walks from $s$ to $\hat{u}$ and from $u$ to $t$, or there exist such walks from $s$ to $u$ and from $\hat{u}$ to $t$.*

*Proof.* For convenience, let us call a walk *special* if it does not pass through $s$ or $t$. Consider a vertex $u \in U(s,t)$. Since $\langle s,t\rangle$ is a panbubble, $U(s,t) = U(t,s)$, which means $u$ is reachable from $s$ without passing through $s$ or $t$, and $u$ is reachable from $t$ without passing through $s$ or $t$. If $u = s$ or $u = t$, the lemma holds trivially. Let us consider $u \in B(s,t)$. Since $u \in B(s,t) = B(t,s)$, there exist special walks from $s$ to $u$ or $\hat{u}$, and there exist special walks from $t$ to $u$ or $\hat{u}$. There are four possible combinations:

(1) There exist special walks from $s$ to $u$ and from $t$ to $u$.
(2) There exist special walks from $s$ to $\hat{u}$ and from $t$ to $\hat{u}$.
(3) There exist special walks from $s$ to $u$ and from $t$ to $\hat{u}$.
(4) There exist special walks from $s$ to $\hat{u}$ and from $t$ to $u$.

At least one of the above must be true. If (3) or (4) holds, then the lemma follows directly. Next, we will prove that if (2) holds, then (3) or (4) must also hold.

Assume for contradiction that (2) holds, while neither (3) nor (4) holds. This means there exist special walks from $s$ to $\hat{u}$ and from $t$ to $\hat{u}$. Since (3) and (4) do not hold, there is no special walk from $s$ to $u$ or from $t$ to $u$. From the contiguity condition of $\langle s,t\rangle$, there exists a walk $\omega$ from $s$ to $t$ in which vertices $u$ and $\hat{u}$ appear. One of the following two sub-cases must hold: Either (a) vertex $u$ appears before the first occurrence of $\hat{u}$ in $\omega$, or (b) vertex $u$ appears after the first occurrence of $\hat{u}$ in $\omega$.

If case (2) and its sub-case (a) hold, then there also exists an alternative walk $\omega'$ from $s$ to $t$ passing through $u$ such that the sub-walk of $\omega'$ from the first vertex $s$ to the first occurrence of $\hat{u}$ is a special walk. Since there is no special walk from $u$ to $t$, the sub-walk of $\omega'$ from the first occurrence of $u$ to the last vertex $t$ must pass through $s$ or $t$. Starting from $u$, the first such ordered pair encountered must be one of $(s,\hat{s}), (\hat{s},s), (t,\hat{t})$ or $(\hat{t},t)$. If $(s,\hat{s})$ or $(t,\hat{t})$ is encountered first, then we would contradict the separable condition of $\langle s,t\rangle$. If $(\hat{s},s)$ is encountered first, then a special walk exists from $s$ to $u$, contradicting our earlier assumption. Similarly, if $(\hat{t},t)$ is encountered first, then a special walk exists from $t$ to $u$, contradicting our earlier assumption.

If case (2) and its sub-case (b) hold, then there also exists an alternative walk from $s$ to $t$ passing through $u$ such that its sub-walk from the first occurrence of $\hat{u}$ to the last vertex $t$ is a special walk. Since there is no special walk from $u$ to $s$, the sub-walk between the first vertex $s$ and the first occurrence of $u$ must pass through $s$ or $t$. One can repeat a similar argument as sub-case (a) to arrive at a contradiction. Thus, if case (2) holds, then case (3) or (4) must hold.

The same argument can be easily extended to show that if (1) holds, then (3) or (4) must also hold. $\square$

# Proof of Lemma 5

**Lemma A.7.** *Consider a panbubble $\langle s,t\rangle$ in $G_b$ such that $\bar{e}_s \prec \bar{e}_t$. For all $u \in U(s,t)$, there exists a root-leaf path containing $\bar{s}$ and $\bar{u}$ in $T_{v_{src}}$.*

*Proof.* Assume for contradiction that there exists a vertex $u \in U(s,t)$ such that there is no root-leaf path containing $\bar{u}$ and $\bar{s}$. Since $u \in U(s,t)$, $u$ is reachable from $s$ without passing through $s$ or $t$ in $G_b$. Accordingly, in $G_U$, there exists a walk with vertex sequence $(\bar{s}, \bar{\hat{s}}, \dots, \bar{u})$ that does not pass through $\bar{s}$ or $\bar{t}$. There also exists a path from $\bar{u}$ to $v_{src}$ that includes neither $\bar{s}$ nor $\bar{t}$. This means $v_{src}$ is reachable from $\bar{\hat{s}}$ without passing through $\bar{s}$ or $\bar{t}$, contradicting our claim in Corollary A.3. $\square$

**Lemma A.8.** *Consider a panbubble $\langle s,t\rangle$ in $G_b$ such that $\bar{e}_s \prec \bar{e}_t$. Let $D_s$ and $D_t$ denote the set of all descendant vertices of edges $\bar{e}_s$ and $\bar{e}_t$, respectively. For all $v \in V_b$, $\bar{v} \in D_s \setminus D_t$ if and only if $v \in B(s,t)$.*

*Proof.* [$\Rightarrow$] Since $\bar{e}_s \prec \bar{e}_t$, we have $\bar{s} \prec \bar{\hat{s}} \prec \bar{\hat{t}} \prec \bar{t}$ (Lemma A.4). If $\bar{v} \in D_s \setminus D_t$, then there exists a path between $\bar{\hat{s}}$ and $\bar{v}$ without passing through $\bar{s}$ or $\bar{t}$. Using Lemma A.2, it follows that $\bar{v} \in \overline{U}(s,t)$. Therefore, $v \in U(s,t)$. Furthermore, $\bar{v}$ cannot be $\bar{s}$ or $\bar{t}$ because $\bar{s}, \bar{t} \notin D_s \setminus D_t$. As a result, $v \in B(s,t)$.

[$\Leftarrow$] Assume for contradiction that $\bar{v} \notin D_s \setminus D_t$. Since $v \in B(s,t)$, (i) there exists a walk $\omega$ with vertex sequence $(\bar{s}, \bar{\hat{s}}, \ldots, \bar{v})$ in $G_U$ which does not pass through $\bar{s}$ or $\bar{t}$, and (ii) there exists a root-leaf path in $T_{v_{src}}$ containing $\bar{s}$ and $\bar{v}$ (Lemma A.7). Since $\bar{v} \notin D_s \setminus D_t$ and $\bar{v}$ is neither $\bar{s}$ nor $\bar{t}$, $\bar{v}$ must be either an ancestor of $\bar{s}$ or a descendant of $\bar{t}$. Since walk $\omega$ from $\bar{s}$ to $\bar{v}$ does not pass through $\bar{s}$ or $\bar{t}$, the bracket sets of $\bar{e}_s$ and $\bar{e}_t$ must be unequal. This contradicts our claim in Lemma A.6. $\square$

**Corollary A.5.** *Consider a panbubble $\langle s,t \rangle$ in $G_b$ such that $\bar{e}_s \prec \bar{e}_t$. For all $u \in B(s,t)$, $\bar{s} \prec \bar{u}$ in $G_U$.*

The above lemmas together form the basis for Lemma 5.

**Lemma 5.** *In a panbubble $\langle s,t \rangle$ in $G_b$, there does not exist any vertex $v \in B(s,t) \setminus \{\hat{s}, \hat{t}\}$ such that, in $G_U$, (a) edge $\bar{e}_v$ has an empty bracket set or contains only a single back edge $\{\bar{v}, \bar{\hat{v}}\}$ and (b) neither $\bar{e}_s \prec \bar{e}_v \prec \bar{e}_t$ nor $\bar{e}_t \prec \bar{e}_v \prec \bar{e}_s$.*

*Proof.* Proof by contradiction. Suppose there exists a vertex $v \in B(s,t) \setminus \{\hat{s}, \hat{t}\}$ such that, in $G_U$: (a) $\bar{e}_v$ has an empty bracket set or contains only a single back edge $\{\bar{v}, \bar{\hat{v}}\}$, and (b) neither $\bar{e}_s \prec \bar{e}_v \prec \bar{e}_t$ nor $\bar{e}_t \prec \bar{e}_v \prec \bar{e}_s$ holds. We will show that the existence of such a vertex $v$ implies the presence of a hairpin entrance vertex in $B(s,t) \setminus \{\hat{s}, \hat{t}\}$, thereby violating the no-hairpin condition of $\langle s,t \rangle$.

Suppose wlog that $\bar{v} \prec \bar{\hat{v}}$. From Lemma A.3, either $\bar{e}_s \prec \bar{e}_t$ or $\bar{e}_t \prec \bar{e}_s$. Assume wlog that $\bar{e}_s \prec \bar{e}_t$. Let $D_s, D_v$ and $D_t$ denote the set of all descendant vertices of edges $\bar{e}_s, \bar{e}_v$, and $\bar{e}_t$ respectively. Since $v \in B(s,t) \setminus \{\hat{s}, \hat{t}\}$, it follows that $\bar{e}_s \prec \bar{e}_v$ and $\bar{v} \in D_s \setminus D_t$ (using Corollary A.5, Lemma A.7). Since $\bar{e}_s \prec \bar{e}_v \prec \bar{e}_t$ does not hold, edges $\bar{e}_v$ and $\bar{e}_t$ do not share a root-leaf path, which means $D_v \subseteq \overline{B}(s,t) \setminus \{\bar{\hat{s}}, \bar{\hat{t}}, \bar{v}\}$ (Lemma A.7).

Since the bracket set of $\bar{e}_v$ is either empty or contains only a single back edge connecting $\{\bar{v}, \bar{\hat{v}}\}$, removing all edges connecting $\bar{v}$ and $\bar{\hat{v}}$ would disconnect the vertices in $D_v$ from the rest of $G_U$. The rest of $G_U$ includes $\bar{s}$ and $\bar{t}$. The same property should hold in $G_b$, that is, the removal of edges connecting $v$ and $\hat{v}$ would disconnect the subgraph induced by vertex set $\{u | u \in V_b, \bar{u} \in D_v\}$ from the rest of the graph (which includes $s$ and $t$). For every vertex $u$ in this vertex set, the contiguity condition of $\langle s,t \rangle$ ensures the existence of a walk from $s$ to $t$ passing through $u$. Thus, the vertex set $\{u | u \in V_b, \bar{u} \in D_v\}$ equals $Y(v) \setminus \{v\}$ and every vertex in $Y(v)$ must appear in an inverted closed walk starting from $v$. As a result, there must exist a hairpin with entrance vertex $v$ or with some other entrance vertex in $Y(v)$. Since $Y(v) \subseteq (B(s,t) \setminus \{\hat{s}, \hat{t}\})$, the no-hairpin condition of $\langle s,t \rangle$ is violated. $\square$

# Proof of Theorem 1

For a hairpin $\langle s \rangle$ in $G_b$, let $\overline{Y}(s)$ denote the set of vertices in $G_U$ corresponding to the set $Y(s)$ in $G_b$. Formally, $\overline{Y}(s) = \{\bar{v} | v \in Y(s)\}$.

**Lemma A.9.** *For each hairpin $\langle s \rangle$ in $G_b$, the set of vertices in $G_U$ reachable from $\bar{\hat{s}}$ without passing through $\bar{s}$ equals $\overline{Y}(s)$.*

*Proof.* We follow the same proof technique used for Lemma A.2. Let $X$ denote the set of vertices in $G_U$ that are reachable from $\bar{\hat{s}}$ without passing through $\bar{s}$. We need to show that $X = \overline{Y}(s)$. We first argue that $\overline{Y}(s) \subseteq X$. Recall that for every $v \in Y(s)$, $v$ is reachable from $s$ without passing through $s$ (from the definition of $Y(s)$). The corresponding walk in $G_U$ contains a subwalk from $\bar{\hat{s}}$ to $\bar{v}$ that does not pass through $\bar{s}$. Therefore, $\bar{v} \in X$.

Next, we prove $X \subseteq \overline{Y}(s)$ by contradiction. Suppose that $X \not\subseteq \overline{Y}(s)$. This implies the existence of a vertex in $G_U$ such that (i) this vertex is reachable from $\bar{\hat{s}}$ without passing through $\bar{s}$, and (ii) this vertex is not in $\overline{Y}(s)$. In that case, there must exist at least one walk $\omega$ having vertex sequence $(\bar{\hat{s}}, v_1, v_2, \ldots, v_k)$ in $G_U$ with $k \geq 1$ such that $\bar{s}$ does not appear in $\omega$, and $v_k$ is the only vertex in $\omega$ that does not belong to $\overline{Y}(s)$. It follows that vertices $v_1, v_2, \ldots, v_k$ are in $\overline{Y}(s) \setminus \{\bar{s}\}$. Further note that $v_k \neq v_{src}$, because if $v_k = v_{src}$, then we would have a tip vertex in $Y(s) \setminus \{s\}$. But such a tip vertex cannot appear in an inverted closed walk starting from $s$, contradicting that $\langle s \rangle$ is a hairpin.

Consider the case when $v_k \neq v_{src}$ and $k > 1$. Since $v_{k-1} \in \overline{Y}(s) \setminus \{\bar{s}\}$ and $v_k \neq v_{src}$, we must have vertices corresponding to $v_{k-1}$ and $v_k$ in $G_b$. We must also have an edge in $G_b$ between $v_{k-1}$ and $v_k$. Since $v_{k-1} \in \overline{Y}(s) \setminus \{\bar{s}\}$ and $v_k \notin \overline{Y}(s)$, observe that the presence of an edge in $G_b$ between a vertex in $Y(s) \setminus \{s\}$ and a vertex in $V_b \setminus Y(s)$ contradicts the separable property of hairpin $\langle s \rangle$.

The other case, i.e., $v_k \neq v_{src}$ and $k = 1$, implies the presence of an edge in $G_b$ between $\hat{s}$ and a vertex in $V_b \setminus Y(s)$, again contradicting the separable property of hairpin $\langle s \rangle$. □

**Corollary A.6.** *For each hairpin $\langle s \rangle$ in $G_b$, the set of vertices in $G_U$ reachable from $\overline{\hat{s}}$ without passing through $\bar{s}$ does not contain $v_{src}$.*

Corollary A.6 further implies the following.

**Corollary A.7.** *For each hairpin $\langle s \rangle$ in $G_b$, $\bar{s} \prec \overline{\hat{s}}$ in $G_U$.*

**Lemma A.10.** *For each hairpin $\langle s \rangle$ in $G_b$, the bracket set of edge $\bar{e}_s$ in $G_U$ is either empty or contains only a single back edge $\{\bar{s}, \overline{\hat{s}}\}$.*

*Proof.* Proof by contradiction. Assume that the bracket set of edge $\bar{e}_s$ contains a back edge other than $\{\bar{s}, \overline{\hat{s}}\}$. From Corollary A.7, we have $\bar{s} \prec \overline{\hat{s}}$ in $G_U$. Let $D_s$ denote the set of all descendant vertices of edge $\bar{e}_s$. Let $A_s$ denote the set of all ancestor vertices of edge $\bar{e}_s$. Based on our assumption on the bracket set of edge $\bar{e}_s$, there must exist a back edge other than $\{\bar{s}, \overline{\hat{s}}\}$ such that one endpoint of the edge is in $D_s$ and the other endpoint of the edge is in $A_s$. By Corollary A.6, the endpoint of this edge in $A_s$ cannot be $v_{src}$. Now, observe that using this edge, we have a path between $\overline{\hat{s}}$ and $\bar{s}$ in $G_U$ without using an edge connecting vertices $\overline{\hat{s}}$ and $\bar{s}$. Mapping this path back to the corresponding sequence of edges in $G_b$ contradicts the separable property of hairpin $\langle s \rangle$. □

**Lemma A.11.** *Consider a hairpin $\langle s \rangle$ in $G_b$. For all $u \in Y(s)$, there exists a root-leaf path containing $\bar{s}$ and $\bar{u}$ in $T_{v_{src}}$.*

*Proof.* Assume for contradiction that there exists a vertex $u \in Y(s)$ such that there is no root-leaf path containing $\bar{u}$ and $\bar{s}$. Since $u \in Y(s)$, $u$ is reachable from $s$ without passing through $s$ in $G_b$. Accordingly, in $G_U$, there exists a walk with vertex sequence $(\bar{s}, \overline{\hat{s}}, \ldots, \bar{u})$ that does not pass through $\bar{s}$. There also exists a path from $\bar{u}$ to $v_{src}$ that does not include $\bar{s}$. This means $v_{src}$ is reachable from $\overline{\hat{s}}$ without passing through $\bar{s}$, contradicting our claim in Corollary A.6. □

**Lemma A.12.** *Consider a hairpin $\langle s \rangle$ in $G_b$. Let $D_s$ denote the set of all descendant vertices of edge $\bar{e}_s$. For all $v \in V_b$, $\bar{v} \in D_s$ if and only if $v \in Y(s) \setminus \{s\}$.*

*Proof.* [$\Rightarrow$] From Corollary A.7, we know $\bar{s} \prec \overline{\hat{s}}$ in $G_U$. If $\bar{v} \in D_s$, then there exists a path between $\overline{\hat{s}}$ and $\bar{v}$ without passing through $\bar{s}$. Using Lemma A.9, it follows that $\bar{v} \in \overline{Y}(s)$. Therefore, $v \in Y(s)$. Furthermore, $\bar{v}$ cannot be $\bar{s}$ because $\bar{s} \notin D_s$. As a result, $v \in Y(s) \setminus \{s\}$.

[$\Leftarrow$] Assume for contradiction that there exists a vertex $v \in Y(s) \setminus \{s\}$ such that $\bar{v} \notin D_s$. Since $v \in Y(s) \setminus \{s\}$, (i) there exists a walk $\omega$ with vertex sequence $(\bar{s}, \overline{\hat{s}}, \ldots, \bar{v})$ in $G_U$ which does not pass through $\bar{s}$, and (ii) there exists a root-leaf path in $T_{v_{src}}$ containing $\bar{s}$ and $\bar{v}$ (Lemma A.11). Since $\bar{v} \notin D_s$ and $\bar{v} \neq \bar{s}$, $\bar{v}$ must be an ancestor of $\bar{s}$. Since walk $\omega$ from $\bar{s}$ to its ancestor vertex $\bar{v}$ does not pass through $\bar{s}$, the bracket set of edge $\bar{e}_s$ must contain a back edge other than $\{\bar{s}, \overline{\hat{s}}\}$. This contradicts our claim in Lemma A.10. □

**Corollary A.8.** *Consider a hairpin $\langle s \rangle$ in $G_b$. For all $u \in Y(s) \setminus \{s\}$, $\bar{s} \prec \bar{u}$ in $G_U$.*

With this, we have all the pieces required to prove Theorem 1.

**Theorem 1** *A vertex pair $s, t \in V_b$ satisfies all the panbubble conditions (excluding minimality) if and only if all the following conditions hold:*

*P1. Either $\bar{s} \prec \bar{\hat{s}} \prec \bar{\hat{t}} \prec \bar{t}$ or $\bar{t} \prec \bar{\hat{t}} \prec \bar{\hat{s}} \prec \bar{s}$ in $G_U$,*
*P2. Edges $\bar{e}_s$ and $\bar{e}_t$ are cycle equivalent in $G_U$,*
*P3. Consider walks in $G_b$ that do not pass through $s$ or $t$. For all vertices $u \in U(s,t) \cup U(t,s)$, there exist such walks from $s$ to $\hat{u}$ and from $u$ to $t$, or there exist such walks from $s$ to $u$ and from $\hat{u}$ to $t$.*
*P4. There is no vertex $v \in B(s,t) \setminus \{\hat{s}, \hat{t}\}$ such that, in $G_U$, (a) edge $\bar{e}_v$ has an empty bracket set or contains only a single back edge $\{\bar{v}, \bar{\hat{v}}\}$ and (b) neither $\bar{e}_s \prec \bar{e}_v \prec \bar{e}_t$ nor $\bar{e}_t \prec \bar{e}_v \prec \bar{e}_s$.*

*Proof.* [$\Rightarrow$] If $\langle s,t \rangle$ is a panbubble in $G_b$, then conditions P1-P4 are directly ensured by Lemma 2, Lemma 3, Lemma 4, and Lemma 5 respectively.

[$\Leftarrow$] For convenience, let us call a walk *special* if it does not pass through $s$ or $t$. From the first condition P1, assume wlog that $\bar{s} \prec \bar{\hat{s}} \prec \bar{\hat{t}} \prec \bar{t}$ in $G_U$. Observe that if every vertex $u \in B(s,t) \cup B(t,s)$ satisfies condition P3, then there exist special walks from $s$ to $t$ passing through $u$, and there exist special walks from $t$ to $s$ passing through $u$. Therefore, both the contiguity and matching conditions are satisfied. This also means $B(s,t) = B(t,s)$.

Next, we prove that the vertex pair $s, t$ satisfies the separable condition. Assume for contradiction that there exist vertices $v_{in} \in B(s,t)$ and $v_{out} \in V_b \setminus B(s,t)$ that remain connected after removing $e_s$ and $e_t$. If $v_{in}$ and $v_{out}$ are connected by a black edge, then both vertices would lie entirely within $B(s,t)$ or entirely outside it. Therefore, $v_{in}$ and $v_{out}$ must be connected by a gray edge.

From condition P3, suppose wlog that there exist special walks from $s$ to $\hat{v}_{in}$ and from $v_{in}$ to $t$. Vertex $v_{out}$ must be either $s$ or $t$; if not, there would exist special walks from $t$ to $\hat{v}_{out}$ passing through $v_{in}$ and $v_{out}$, which would imply $v_{out} \in B(s,t)$. Therefore, $v_{out} \in \{s,t\}$. Now observe that (i) there exist walks starting from vertex $\bar{s}$ and edge $\bar{e}_s$ to $\bar{v}_{in}$ in $G_U$ without passing through $\bar{t}$, and (ii) there exist walks starting from vertex $\bar{t}$ and edge $\bar{e}_t$ to $\bar{v}_{in}$ in $G_U$ without passing through $\bar{s}$. An edge from $\bar{v}_{in}$ to either $\bar{s}$ or $\bar{t}$ would break the cycle-equivalence of edges $\bar{e}_s$ and $\bar{e}_t$, violating condition P2.

So far, we have established that the vertex pair $s, t$ satisfies the matching, contiguity, and the separable condition. Finally, we show that the vertex pair $s, t$ satisfies the no-hairpin condition. Assume, for contradiction, that there exists a vertex $v \in B(s,t) \setminus \{\hat{s}, \hat{t}\}$ that serves as an entrance vertex of a hairpin. By Corollary A.7, it follows that $\bar{v} \prec \bar{\hat{v}}$ in $G_U$. From Lemma A.10, the bracket set of edge $\bar{e}_v$ is either empty or only contains the single back edge $\{\bar{v}, \bar{\hat{v}}\}$. Since our proof of Lemma A.8 depends solely on the matching, contiguity, and separable conditions, we have $\bar{v} \in D_s \setminus D_t$, where $D_s$ and $D_t$ denote the set of all descendant vertices of edges $\bar{e}_s$ and $\bar{e}_s$, respectively. This observation, along with condition P4, implies that the ordering $\bar{e}_s \prec \bar{e}_v \prec \bar{e}_t$ must hold. This means $\bar{t} \in D_v$, where $D_v$ is the set of all descendant vertices of edge $\bar{e}_v$, and hence $t \in Y(v) \setminus \{v\}$ (by Lemma A.12). It follows that $t$ appears in some inverted closed walk in $G_b$ starting from $v$ while $s$ does not. Therefore, there exists a cycle in $G_U$ containing $\bar{e}_t$ but not $\bar{e}_s$, which contradicts the cycle-equivalence of edges $\bar{e}_s$ and $\bar{e}_t$ (condition P2). □

## Proof of Theorem 2

**Lemma A.13.** *In a hairpin $\langle s \rangle$ in $G_b$, there does not exist any vertex $v \in Y(s) \setminus \{s, \hat{s}\}$ such that the bracket set of edge $\bar{e}_v$ in $G_U$ is either empty or contains only a single back edge $\{\bar{v}, \bar{\hat{v}}\}$.*

*Proof.* Proof by contradiction. Suppose there exists a vertex $v \in Y(s) \setminus \{s, \hat{s}\}$ such that the bracket set of $\bar{e}_v$ in $G_U$ is either empty or contains only a single back edge $\{\bar{v}, \bar{\hat{v}}\}$. Suppose wlog that $\bar{v} \prec \bar{\hat{v}}$. From Corollary A.7, we have $\bar{s} \prec \bar{\hat{s}}$. From Corollary A.8, we have $\bar{s} \prec \bar{v}$. Therefore, $\bar{s} \prec \bar{\hat{s}} \prec \bar{v} \prec \bar{\hat{v}}$. Let $D_v$ denote the set of all descendant vertices of edges $\bar{e}_v$. From Lemma A.12, $D_v \subset \overline{Y}(s)$. Now, from the definition of hairpin, each vertex $u \in V_b$ such that $\bar{u} \in D_v$ must appear in an inverted closed walk in $G_b$ starting from $s$. Since the bracket set of $\bar{e}_v$ in $G_U$ is either empty or or contains only a single back edge $\{\bar{v}, \bar{\hat{v}}\}$, such an inverted closed walk must include another nested inverted closed walk inside it that starts from $v$ and contains $u$. This implies each vertex in $Y(v)$ appears in an inverted closed walk starting from $v$. Furthermore, removing all edges between vertices $v$ and $\hat{v}$ would disconnect $Y(v) \setminus \{v\}$ from the remaining graph. This means vertex

$v$ also satisfies the hairpin conditions. But then $\langle s \rangle$ cannot be a hairpin by the hairpin definition. We arrive at a contradiction. $\square$

**Lemma A.14.** *Suppose $\langle s \rangle$ is a hairpin in $G_b$. Consider walks in $G_b$ that do not pass through $s$. For every vertex $u \in Y(s)$, there exist such walks from $s$ to $u$ and from $s$ to $\hat{u}$.*

*Proof.* We follow the same proof technique used for Lemma 4. Because $\hat{s} \in Y(s)$ appears on an inverted closed walk starting from $s$, the claim holds trivially for $u = s$. Therefore, consider a vertex $u \in Y(s) \setminus \{s\}$. For convenience, let us call a walk *special* if it does not pass through $s$. Since vertex $u$ belongs to set $Y(s) \setminus \{s\}$, $u$ is reachable from $s$ without passing through $s$. Accordingly, there exist special walks from $s$ to $u$ or $\hat{u}$. If there exist special walks from $s$ to both $u$ and $\hat{u}$, then the lemma follows. The other possibility is that there exist special walks from $s$ to $u$ but not to $\hat{u}$ (or vice-versa). We will argue that this is not possible.

Assume for contradiction that special walks from $s$ to $u$ exist, but no special walk from $s$ to $\hat{u}$ exists. Since $\langle s \rangle$ is a hairpin and vertex $u \in Y(s) \setminus \{s\}$, $u$ lies on an inverted closed walk $\omega$ starting from $s$. We consider two cases depending on whether $\hat{u}$ or $u$ appears first in $\omega$.

Case 1 ($\hat{u}$ appears before the first occurrence of $u$ in $\omega$): Because a special walk from $s$ to $u$ exists, there also exists an alternative inverted closed walk $\omega'$ from $s$ containing $u$ in which the sub-walk from the first vertex $s$ to the first occurrence of $u$ is special. Since there is no special walk from $\hat{u}$ to $s$, the sub-walk of $\omega'$ from the first occurrence of $\hat{u}$ to the last vertex $s$ must pass through $s$. Starting from $\hat{u}$, the first such ordered pair encountered must be one of $(s, \hat{s})$ or $(\hat{s}, s)$. If it is $(s, \hat{s})$, then we would contradict the separable property (Condition (ii)) of hairpin $\langle s \rangle$. If it is $(\hat{s}, s)$, then we have a special walk from $\hat{u}$ to $s$, contradicting our earlier assumption.

Case 2 ($\hat{u}$ appears after the first occurrence of $u$ in $\omega$). Because a special walk from $s$ to $u$ exists, there also exists an alternative inverted closed walk from $s$ containing $u$ in which the sub-walk from the first occurrence of $u$ to the last vertex $s$ is special. Since there is no special walk between $\hat{u}$ and $s$, the sub-walk of $\omega'$ between the first vertex $s$ and the first occurrence of $\hat{u}$ must pass through $s$. A similar line of reasoning to that in Case 1 results in a contradiction. $\square$

We now have all the lemmas necessary to prove Theorem 2.

**Theorem 2** *A vertex $s$ is an entrance vertex of a hairpin in $G_b$ if and only if all the following hold:*

- *H1. The bracket set of edge $\bar{e}_s$ in $G_U$ is either empty or contains a single back edge connecting $\bar{s}$ and $\bar{\hat{s}}$.*
- *H2. There does not exist any vertex $v \in V_b \setminus \{\hat{s}\}$ which satisfies the above condition while also satisfying $\bar{s} \prec \bar{v}$ in $G_U$.*
- *H3. Consider walks in $G_b$ that do not pass through $s$. For all vertices $u \in Y(s)$, there exist such walks from $s$ to $u$ and from $s$ to $\hat{u}$.*

*Proof.* [$\Rightarrow$] If vertex $s \in V_b$ is an entrance vertex of a hairpin in $G_b$, then all conditions H1, H2, and H3 directly hold due to Lemma A.10, Lemmas A.12-A.13, and Lemma A.14, respectively.

[$\Leftarrow$] For convenience, let us call a walk *special* if it does not pass through $s$. By condition H3, there exist special walks from $s$ to $s$ in which vertex $u$ appears for all $u \in Y(s)$. In other words, $\forall u \in Y(s)$, $u$ belongs to an inverted closed walk starting from $s$. Hence, the first condition in the hairpin definition is satisfied. Next, we will argue that the second condition holds.

Assume for contradiction that there exist vertices $v_{in} \in Y(s) \setminus \{s\}$ and $v_{out} \in V_b \setminus (Y(s) \setminus \{s\})$ which remain connected after removing all the edges between $s$ and $\hat{s}$. If $v_{in}$ and $v_{out}$ are connected by a black edge, then both vertices would entirely belong to one partition, either within $Y(s) \setminus \{s\}$ or its complement. Therefore, $v_{in}$ and $v_{out}$ must be connected by a gray edge.

Using condition H3, there exist special walks from $s$ to $\hat{v}_{in}$ and from $v_{in}$ to $s$. Vertex $v_{out}$ must be equal to $s$; if not, there would exist special walks from $t$ to $\hat{v}_{out}$ passing through $v_{in}$ and $v_{out}$, which would imply that $v_{out} \in Y(s) \setminus \{s\}$. Therefore, $v_{out} = s$. Furthermore, $v_{in}$ cannot be $\hat{s}$ because vertices $v_{in}$ and $v_{out} = s$ are assumed to remain connected after removing all edges between $s$ and $\hat{s}$. Now observe that there exist walks starting from vertex $\bar{s}$ and edge $\bar{e}_s$ to $\bar{v}_{in}$ in $G_U$ without passing through $\bar{s}$. An edge between $\bar{v}_{in}$ and

$\bar{v}_{out} = \bar{s}$ implies that vertex $\bar{s}$ is part of a cycle in $G_U$ that includes vertex $\bar{v}_{in}$. However, this is not possible because condition H1 ensures that the bracket set of edge $\bar{e}_s$ in $G_U$ is either empty or contains only a single back edge $\{\bar{s}, \hat{\bar{s}}\}$.

So far, we have established that vertex $s$ satisfies the first two conditions of the hairpin definition. To complete the proof, we need to argue that no vertex in $Y(s)$ other than $s$ satisfies the first two conditions from the hairpin definition. Assume, for contradiction, that there exists a vertex $v \in Y(s) \setminus \{s\}$ that satisfies the conditions. Because our proofs of Corollary A.8 and Lemma A.10 relied only on the first two conditions from the hairpin definition, we conclude that $\bar{s} \prec \bar{v}$ in $G_U$, and the bracket set of edge $\bar{e}_v$ is either empty or contains only a single back edge $\{\bar{v}, \hat{\bar{v}}\}$. But this contradicts condition H2. Hence, $\langle s \rangle$ is a hairpin. $\square$

# Proof of Theorem 3

**Lemma A.15.** *Any vertex can be the entrance of at most one panbubble.*

*Proof.* Proof by contradiction. Suppose there exist two distinct panbubbles $\langle s, t_1 \rangle$, $\langle s, t_2 \rangle$, $t_1 \neq t_2$. If $t_2 \in U(s, t_1)$, then $t_2$ is also in $B(s, t_1)$. However, this contradicts the minimality condition for $\langle s, t_1 \rangle$. Similarly, $t_1$ being a vertex in $U(s, t_2)$ also results in a contradiction.

Now suppose that $t_2 \notin U(s, t_1)$. Since $\langle s, t_2 \rangle$ is a panbubble, $t_2$ is reachable from $s$ using some walk that does not pass through $s$ or $t_2$ (from the definition of $U(s, t_2)$). This means there exists a walk $\omega$ between $s$ and $t_2$ in which both $s$ and $t_2$ appear only once. Because $t_2$ appears in $\omega$, it would be correct to state that not all vertices in walk $\omega$ belong to $U(s, t_1)$. Consider the ordered list of the vertices in walk $\omega$ starting from vertex $s$. Suppose $v$ is be first vertex in $\omega$ that is not in $U(s, t_1)$. The vertex just before $v$ must be either $s$ or $t_1$. Since $s$ does not appear twice in $\omega$, this vertex must be $t_1$. But the appearance of $t_1$ in $\omega$ implies that $t_1$ is in $U(s, t_2)$, which violates the first half of the argument. $\square$

**Theorem 3** *Let $\langle x, y \rangle$ and $\langle m, n \rangle$ be two different panbubbles such that $B(x, y) \cap B(m, n) \neq \phi$, then either $B(x, y)$ is a strict subset of $B(m, n)$ or $B(m, n)$ is a strict subset of $B(x, y)$.*

*Proof.* Since $\langle x, y \rangle$ and $\langle m, n \rangle$ be two different panbubbles, $x \neq y \neq m \neq n$ (using Lemma A.15). Suppose wlog that $\bar{e}_x \prec \bar{e}_y$ and $\bar{e}_m \prec \bar{e}_n$ (and hence by Lemma A.4, we have $\bar{x} \prec \hat{\bar{x}} \prec \hat{\bar{y}} \prec \bar{y}$ and $\bar{m} \prec \hat{\bar{m}} \prec \hat{\bar{n}} \prec \bar{n}$ respectively). Consider any vertex $v$ in $B(x, y) \cap B(m, n)$. From Corollary A.5, it follows that $\bar{x} \prec \bar{v}$ and $\bar{m} \prec \bar{v}$. Therefore, $\bar{x}$ and $\bar{m}$ share a root-leaf path, which means either $\bar{x} \prec \bar{m}$ or $\bar{m} \prec \bar{x}$. Let us consider these two cases separately.

Case 1 ($\bar{x} \prec \bar{m}$): In this case, we will prove that $B(m, n) \subset B(x, y)$. Assume for contradiction that $B(m, n) \not\subset B(x, y)$. Since $\bar{x} \prec \bar{m}$, we have $\bar{x} \prec \hat{\bar{x}} \prec \bar{m} \prec \hat{\bar{m}}$ and $\bar{e}_x \prec \bar{e}_m$. Let us use $D_x, D_y$, and $D_m$ to denote the set of all descendant vertices of edges $\bar{e}_x$, $\bar{e}_y$, and $\bar{e}_m$, respectively. Clearly, $\bar{y}, \hat{\bar{y}} \in D_x$. Let us consider the different positions possible for vertex $\bar{y}$ relative to tree edge $\bar{e}_m$ in $T_{v_{src}}$. Either (i) $\bar{x} \prec \bar{y} \prec \bar{m}$, (ii) $\bar{y}$ and $\bar{m}$ do not share any root-leaf path, or (iii) $\bar{m} \prec \bar{y}$ (Figure S1A). Condition (i) cannot hold because $\bar{e}_x \prec \bar{e}_y \prec \bar{e}_m$ implies $B(x, y) \cap B(m, n) = \phi$ (Lemma A.8). If (ii) holds, then all vertices in set $\overline{B}(m, n)$ belong to $D_x \setminus D_y = \overline{B}(x, y)$ (Lemma A.8) and we are done. Henceforth, we will assume that (iii) holds. If $\bar{m} \prec \bar{y} = \hat{\bar{m}}$, then again $B(x, y) \cap B(m, n) = \phi$ (Lemma A.8). Therefore, $\bar{e}_x \prec \bar{e}_m \prec \bar{e}_y$.

Let us also consider the possibilities of the position of vertex $\bar{n}$ relative to tree edge $\bar{e}_y$ in $T_{v_{src}}$. Either (a) $\bar{m} \prec \bar{n} \prec \bar{y}$, (b) $\bar{n}$ and $\bar{y}$ do not share any root-leaf path, or (c) $\bar{y} \prec \bar{n}$ (Figure S1B). If (a) holds, then all vertices in set $\overline{B}(m, n)$ belong to $D_x \setminus D_y = \overline{B}(x, y)$ (Lemma A.8) and we are done. Next, we will argue that (b) cannot hold. Assume for contradiction that (b) holds. All vertices in $D_y$ belong to $D_m \setminus D_n = \overline{B}(m, n)$ (Lemma A.8). Observe that $\bar{e}_y$ must have an empty bracket set. This is because $\bar{e}_x$ and $\bar{e}_y$ have identical bracket sets (Lemma A.6), and having an edge between a descendant of $\bar{e}_y$ to an ancestor of $\bar{e}_x$ would contradict the separable condition of $\langle m, n \rangle$. Now, the contiguity condition of $\langle m, n \rangle$ suggests that for all vertices $v \in V_b$ such that $\bar{v} \in D_y$, there exists a walk in $G_b$ from $m$ to $n$ passing through $v$. Since $\bar{e}_y$ has an empty bracket set in $G_U$, this is only possible if there exists a hairpin entrance vertex at $u \in V_b$ such that $\bar{u} \in D_y$. However, this would contradict the no-hairpin condition of $\langle m, n \rangle$.

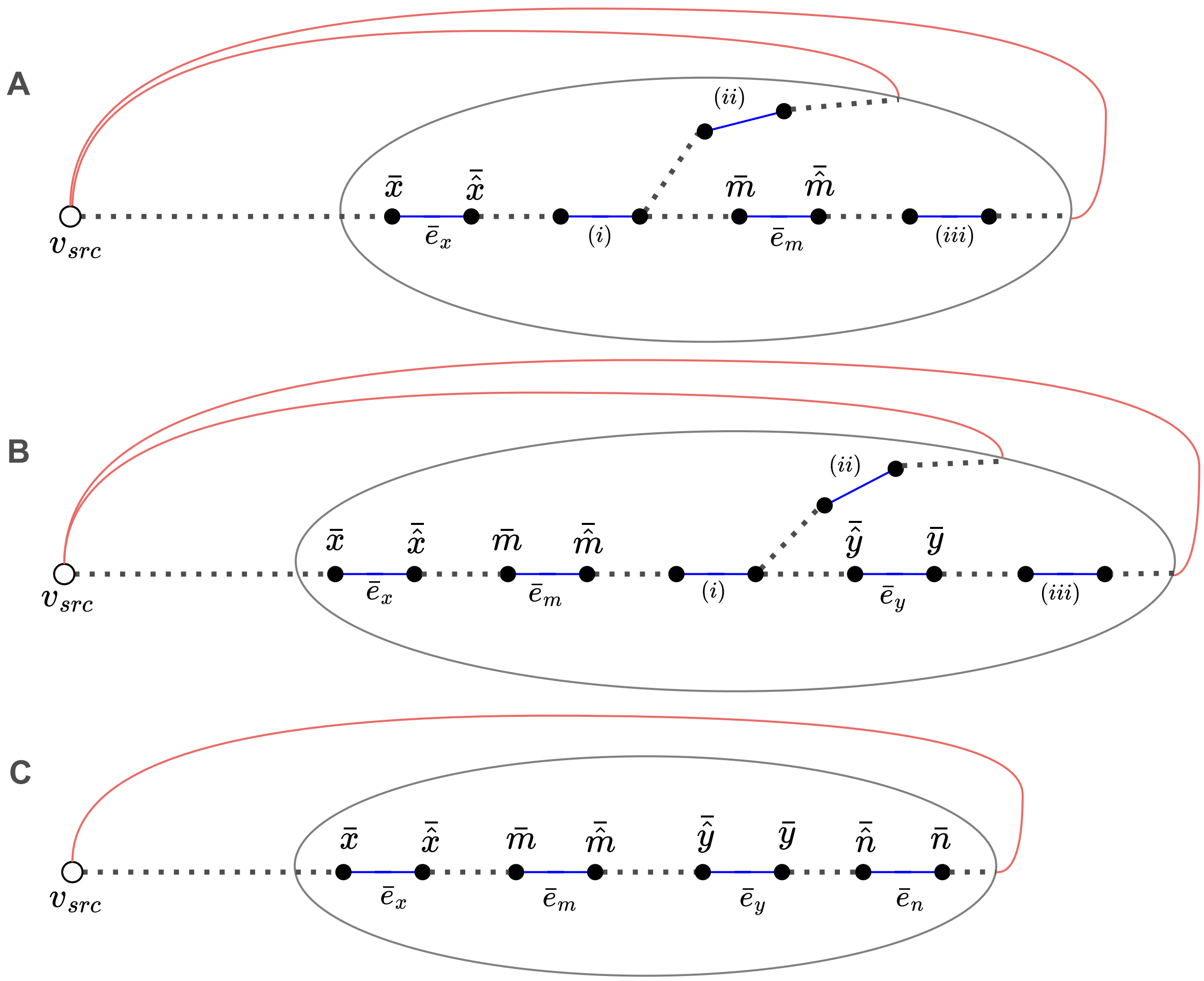


**Figure S1:** Illustration of the case when $\bar{x} \prec \bar{m}$. **(A)** Depicts the three possible positions of vertex $\bar{y}$ relative to tree edges $\bar{e}_x$ and $\bar{e}_m$. **(B)** Shows the three possible positions of vertex $\bar{n}$ relative to tree edges $\bar{e}_m$ and $\bar{e}_y$. **(C)** Presents the configuration satisfying $\bar{e}_x \prec \bar{e}_m \prec \bar{e}_y \prec \bar{e}_n$.

Henceforth, we assume that (c) holds. As a result, $\bar{e}_x \prec \bar{e}_m \prec \bar{e}_y \prec \bar{e}_n$ (Figure S1C). In the remaining part of this proof, we will argue that $\bar{e}_x \prec \bar{e}_m \prec \bar{e}_y \prec \bar{e}_n$ is also not possible by contradiction. We will show that $\langle m, y\rangle$ is a panbubble if $\bar{e}_x \prec \bar{e}_m \prec \bar{e}_y \prec \bar{e}_n$, which would then contradict the minimality condition of $\langle x, y\rangle$ and $\langle m, n\rangle$.

Note that $\bar{e}_x$ and $\bar{e}_y$ have identical bracket sets (Lemma A.6). Similarly, $\bar{e}_m$ and $\bar{e}_n$ have identical bracket sets. Since $\bar{e}_x \prec \bar{e}_m \prec \bar{e}_y \prec \bar{e}_n$, it follows that $\bar{e}_x, \bar{e}_m, \bar{e}_y$, and $\bar{e}_n$ have identical bracket sets. Therefore, $\bar{m}$ or $\bar{y}$ must appear in a walk from a vertex in $D_m \setminus D_y$ to a vertex not in $D_m \setminus D_y$. In the following, we show that the vertex pair $(m, y)$ in $G_b$ satisfies all the panbubble conditions.

- *matching:* Assume for contradiction that $U(m, y) \neq U(y, m)$. Suppose wlog that $U(m, y) \setminus U(y, m) \neq \phi$. Choose any vertex $v$ from $U(m, y) \setminus U(y, m)$. Note that $v$ is reachable from $m$ without passing through $m$ or $y$. However, $v$ is not reachable from $y$ without passing through $m$ or $y$. Since $y, \hat{y} \in U(y, m)$, $v$ is neither $y$ nor $\hat{y}$.
  If $v$ is reachable from $m$ without passing through $m$ or $y$ in $G_b$, then $\bar{v}$ appears in some walk of type $(\bar{m}, \overline{\hat{m}}, \ldots)$ in $G_U$ which does not pass through $\bar{m}$ or $\bar{y}$. Since $\bar{e}_m \prec \bar{e}_y$, and $\bar{e}_m, \bar{e}_y$ have identical bracket sets, it follows that $\bar{v} \in D_m \cup \{\bar{m}\} \setminus D_y$.
  From the above arguments, note that $v$ also belongs to $B(m, n)$ (Lemma A.8). If $\bar{e}_m, \bar{e}_y$ have identical bracket sets in $G_U$, then, in $G_b$, $v$ is reachable from $n$ without passing through $m$ or $n$ if only if $v$ is reachable from $y$ without passing through $m$ or $y$. Since $v \notin U(y, m)$, $v$ is not reachable from $n$ without passing through $m$ or $n$. This contradicts the matching condition of $\langle m, n\rangle$.
- *separable:* From the above arguments, it is clear that every vertex in $\overline{B}(m, y)$ belongs to the set $D_m \setminus D_y$. To prove separability, it suffices to show that $D_m \setminus D_y = \overline{B}(m, y)$. Assume for contradiction that there exists $v \in V_b$ such that $\bar{v} \in D_m \setminus D_y$ and $v$ is not reachable from $m$ without passing through $m$ or $y$. Since $\bar{v} \in D_m \setminus D_y$, $v$ belongs to $B(x, y)$. Thus, $v$ is reachable from $x$ without passing through $x$ or $y$. In other words, there exists a walk from $x$ to $v$ or $\hat{v}$ without passing through $x$ or $y$. Since $\bar{e}_x, \bar{e}_m$ and $\bar{e}_y$ have identical bracket sets in $G_U$, this walk must pass through $m$ followed by $\hat{m}$ in $G_b$. This makes $v$ reachable from $m$ without passing through $m$ or $y$, leading to a contradiction.
- *contiguity:* The above arguments also imply that $\overline{U}(m, y) \subset \overline{U}(x, y)$. Now since $\bar{e}_x, \bar{e}_m$ and $\bar{e}_y$ have identical bracket sets, all walks of type $(\bar{x}, \overline{\hat{x}}, \ldots \overline{\hat{y}}, \bar{y})$ in $G_U$ must include $\bar{m}$ immediately followed by $\overline{\hat{m}}$. In each of these walks, observe that the vertices that are in $D_m \setminus D_y$ must appear after the first appearance of $\bar{m}$. Recall that the contiguity condition of $\langle x, y\rangle$ guarantees a walk in $G_b$ from $x$ to $y$ passing through $v$ for all $v$ in $U(x, y)$ (and hence for all $v$ in $U(m, y)$). It further guarantees a walk from $m$ to $y$ passing through $v$ for all $v \in U(m, y)$.
- *no-hairpin:* Since $U(m, y) \subset U(x, y)$, the presence of a hairpin at any vertex $v$ in $B(m, y) \setminus \{\hat{m}, \hat{y}\}$ would violate the no-hairpin condition of $\langle x, y\rangle$.
- *minimality:* If there is any vertex $v \in U(m, y)$ other than $y$ which pairs with $m$ and satisfies the above criteria, it would contradict the minimality condition of $\langle m, n\rangle$. Similarly, if there is any vertex $v \in U(m, y)$ other than $m$ which pairs with $y$ and satisfies the above criteria, it would contradict the minimality condition of $\langle x, y\rangle$.

Therefore, $\langle m, y\rangle$ is a panbubble. But this contradicts the minimality condition of $\langle x, y\rangle$ and $\langle m, n\rangle$.

Case 2 ($\bar{m} \prec \bar{x}$): By a symmetric argument, we obtain $B(x, y) \subset B(m, n)$. □

# Proof of Theorem 4 and Theorem 5

The following holds due to the separable condition in our panbubble definition.

**Lemma A.16.** *Suppose $\langle s, t\rangle$ is a panbubble in biedged graph $G_b$. There is no gray edge in $G_b$ that connects vertices $s$ and $\hat{s}$, and there is no gray edge in $G_b$ that connects vertices $t$ and $\hat{t}$.*

**Theorem 4** *Let $\langle x, y\rangle$ and $\langle z\rangle$ be a panbubble and a hairpin respectively in $G_b$ such that $B(x, y) \cap (Y(z) \setminus \{z\}) \neq \phi$, then $B(x, y)$ is a strict subset of $Y(z) \setminus \{z\}$.*

*Proof.* Suppose wlog that $\bar{e}_x \prec \bar{e}_y$ in $G_U$. Therefore, $\bar{x} \prec \overline{\hat{x}} \prec \overline{\hat{y}} \prec \bar{y}$ (Lemma A.4). From Corollary A.7, we also have $\bar{z} \prec \overline{\hat{z}}$. Let us use $D_x, D_y$, and $D_z$ to denote the sets of all descendant vertices of edges $\bar{e}_x$, $\bar{e}_y$, and $\bar{e}_z$, respectively. From Lemma A.8 and Lemma A.12, $D_x \setminus D_y = \overline{B}(x, y)$ and $D_z = \overline{Y}(z) \setminus \{\bar{z}\}$.

We consider all possible positions of edge $\bar{e}_z$ relative to edges $\bar{e}_x$ and $\bar{e}_y$: (i) $\bar{e}_x$ and $\bar{e}_z$ do not share a root-leaf path, (ii) $\bar{e}_z \prec \bar{e}_x$, (iii) $\bar{e}_z = \bar{e}_x$, (iv) $\bar{e}_x \prec \bar{e}_z \prec \bar{e}_y$, (v) $\bar{e}_x \prec \bar{e}_z$ with $\bar{e}_x$ and $\bar{e}_y$ not sharing a root-leaf path, (vi) $\bar{e}_z = \bar{e}_y$, and (vii) $\bar{e}_x \prec \bar{e}_y \prec \bar{e}_z$. Below, we will argue that all the cases except (ii) are not possible.

Since panbubble $\langle x, y \rangle$ and hairpin $\langle z \rangle$ share one or more vertices (i.e., $B(x, y) \cap (Y(z) \setminus \{z\}) \neq \phi$), it is easy to observe that cases (i), (vi), and (vii) are not possible. Next, we consider case (iii). Using Lemma A.16 and Lemma A.10, we conclude that edge $\bar{e}_x$ has an empty bracket set. This means $\bar{e}_y$ also has an empty bracket set (Lemma A.6). But this contradicts our earlier claim that the interior of a hairpin cannot contain an edge with an empty bracket set (Lemma A.13). Lastly, cases (iv) and (v) are not possible because they contradict the no-hairpin condition of $\langle x, y \rangle$.

We are only left with case (ii), in which $\bar{e}_z \prec \bar{e}_x \prec \bar{e}_y$. In this case, observe that $D_x \setminus D_y = \overline{B}(x, y)$ is a strict subset of $D_z = \overline{Y}(z) \setminus \{\bar{z}\}$. Hence, the claim follows. ☐

**Theorem 5** *Let $\langle m \rangle$ and $\langle n \rangle$ be two distinct hairpins in $G_b$. Then, the two hairpins do not share any vertex, i.e., $(Y(m) \setminus \{m\}) \cap (Y(n) \setminus \{n\})$ is an empty set.*

*Proof.* From Corollary A.7, we have $\bar{m} \prec \overline{\hat{m}}$ and $\bar{n} \prec \overline{\hat{n}}$ in $G_U$. Since $\langle m \rangle \neq \langle n \rangle$, $m \neq n$ and $e_m \neq e_n$. Either $\bar{e}_m$ and $\bar{e}_n$ share a root leaf path or they do not. In the first case, it is clear that $(\overline{Y}(m) \setminus \{\bar{m}\}) \cap (\overline{Y}(n) \setminus \{\bar{n}\})$ is an empty set (Lemma A.12) and we are done.

Now consider the second case where $\bar{e}_m$ and $\bar{e}_n$ share a root leaf path. We argue that this case is not possible. Suppose wlog that $\bar{e}_m \prec \bar{e}_n$. From Lemma A.10, bracket set of edge $\bar{e}_n$ is either empty or contains only a single back edge $\{\bar{n}, \overline{\hat{n}}\}$. Using Lemma A.9, we know that $\bar{n} \in \overline{Y}(m) \setminus \{\bar{m}, \overline{\hat{m}}\}$. However, this contradicts our earlier claim that such a vertex cannot be present in the interior of a hairpin (Lemma A.13). ☐

# B Algorithms to detect panbubbles and hairpins: full exposition

## Proof of Theorem 6

**Theorem 6** *Given a compact biedged graph $G(V_b, E_b)$, the entrance vertices of all panbubbles and hairpins can be enumerated exactly in $\mathcal{O}(|V_b|^2(|V_b| + |E_b|))$ time. Moreover, the corresponding heuristic algorithm for the same task uses $\mathcal{O}(|V_b|(|V_b| + |E_b|))$ time.*

*Proof.* Let us first analyze the exact algorithm for panbubble detection. Clearly, the initial steps of computing $G_U$, $T_{v_{src}}$, $A_d$, $A_f$, $A_{bridge}$, and the cycle-equivalent classes $C_1, C_2, \ldots, C_k$ require $\mathcal{O}(|V_b| + |E_b|)$ time. Let $b_i$ denote the count of edges in cycle-equivalent class $C_i$ whose corresponding edge in $G_b$ is black. For any vertex pair $(s, t)$, where $s, t \in V_b$, the computation of arrays $W_s, W_t$ can be done in $\mathcal{O}(|V_b| + |E_b|)$ time. By using these two arrays and our precomputed arrays $A_d$, $A_f$, and $A_{bridge}$, all the checks mentioned in Section 4.1 can be made in $\mathcal{O}(1)$ time for every $v \in V_b$. Therefore, the total worst-case runtime to process all cycle-equivalent classes comes out to be $\mathcal{O}(\sum_{i=1}^{k} b_i^2(|V_b| + |E_b|))$ time. Now recall that for every vertex in $G_b$ there is exactly one black edge. Therefore, $\sum_{i=1}^{k} b_i$ is bounded by $|V_b|$, which further implies that $\sum_{i=1}^{k} b_i^2$ is bounded by $|V_b|^2$. Accordingly, the runtime complexity of detecting all panbubbles can be expressed as $\mathcal{O}(|V_b|^2(|V_b| + |E_b|))$. Using a similar argument, it follows that the worst-case time complexity of the heuristic algorithm for detecting panbubbles is $\mathcal{O}(|V_b|(|V_b| + |E_b|))$.

Next, we analyze the runtime of the exact algorithm for detecting hairpins. The computation of array $W$ to verify a hairpin entrance vertex can be done in $\mathcal{O}(|V_b| + |E_b|)$ time. Using array $W$ and the other precomputed arrays, all the checks mentioned in Section 4.2 can be made in $\mathcal{O}(1)$ time for every $v \in V_b$. Therefore, the worst-case time complexity of detecting all hairpins is $\mathcal{O}(|V_b|(|V_b| + |E_b|))$.

Note that both our exact and heuristic implementations compute hairpins exactly; they differ only in the algorithm used for computing panbubbles. ☐

# C Supplementary figures

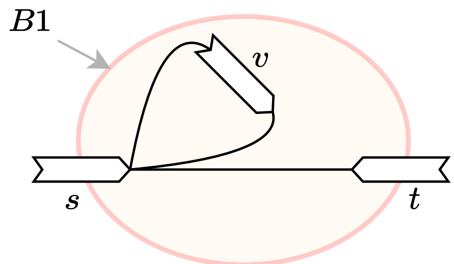


**Figure S2:** In the above bidirected graph[†], bubble $B1$ is both a snarl [24] and a flubble [22] by definition. $B1$ includes vertex $v$ that does not lie on the walk from source vertex $s$ to sink vertex $t$.

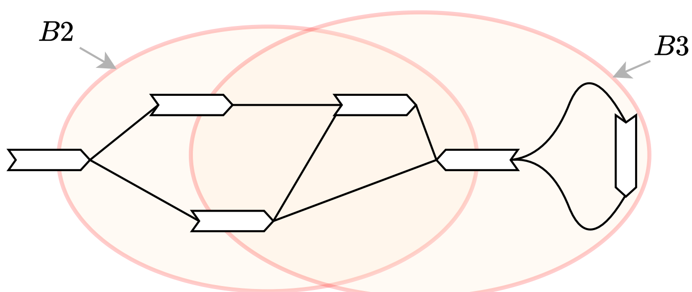


**Figure S3:** Bubbles $B2$ and $B3$ in the above bidirected graph are valid snarls [24], bibubbles [20], and flubbles [22] by definition. Bubbles $B2$ and $B3$ overlap, i.e, the two subgraphs share vertices. VG includes a post-processing step to remove a subset of snarls from the final output, preventing such overlaps [24].

[†]In a bidirected graph, each vertex has two sides, and a valid walk must enter a vertex on one side and exit through the opposite side. Readers seeking a formal introduction to bidirected graphs are referred to [25].

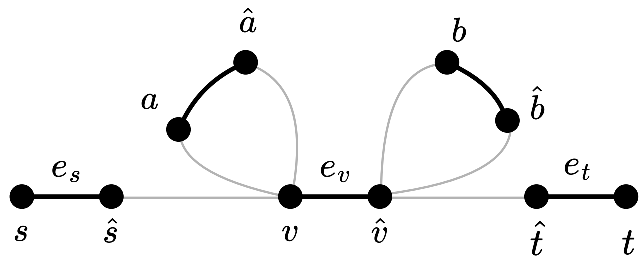


**Figure S4:** An example adversarial case: Panbubble $\langle s, t\rangle$ contains an edge $e_v$ such that $\bar{e}_v, \bar{e}_s$, and $\bar{e}_t$ are cycle equivalent in $G_U$.

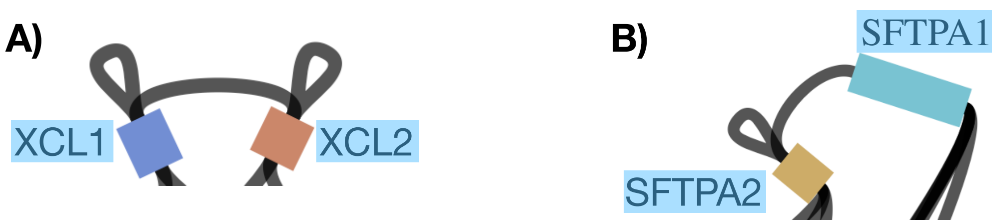


**Figure S5:** Bandage visualization of two subgraphs that are not marked as snarls by VG in graph $G4$.

# D Supplementary tables

| Tool | Version | Commands |
|---|---|---|
| Billi (exact) | v1.0 | billi decompose -i <INPUT_GFA_FILE> -e ><br><OUTPUT_FILE_PATH> |
| Billi (heuristic) | v1.0 | billi decompose -i <INPUT_GFA_FILE> ><br><OUTPUT_FILE_PATH> |
| VG | v1.61.0 | (For graphs $G1 - G7$)<br>vg convert -g <INPUT_GFA_FILE> > FILE.vg<br>vg index -x FILE.xg FILE.vg<br>vg snarls -t 48 FILE.xg > <OUTPUT_FILE_PATH><br><br>(For graph $G8$)<br>vg snarls -t 48 <INPUT_GFA_FILE> > <OUTPUT_FILE_PATH> |
| Pangene | commit:fc3366a | k8 pangene.js call <INPUT_GFA_FILE> ><br><OUTPUT_FILE_PATH> |

**Table S1:** Commands used to evaluate the tools. For VG, only the `vg snarls` command was timed. We used a different command to run VG on graph $G8$ as suggested by VG developers[‡].

| Graph | Panbubble | | | Hairpin | | |
|---|---|---|---|---|---|---|
| | Min | Median | Max | Min | Median | Max |
| $G1$ | 6 | 6 | 28 | – | – | – |
| $G2$ | 4 | 6 | 496 | – | – | – |
| $G3$ | 4 | 4 | 60 | 3 | 3 | 3 |
| $G4$ | 2 | 4 | 58 | 1 | 1 | 3 |
| $G5$ | 4 | 6 | 282,690 | – | – | – |
| $G6$ | 4 | 6 | 163,132 | 3 | 3 | 11 |
| $G7$ | 4 | 6 | 1,436,456 | 747 | 747 | 747 |
| $G8$ | 4 | 6 | 2,293,044 | 3 | 11 | 2,259 |

**Table S2:** Size statistics for all panbubbles and hairpins identified by Billi in pangenome graphs $G1$–$G8$. For each panbubble $\langle s, t\rangle$, we report its size as the number of vertices in its subgraph (i.e., $|B(s,t)|$). Likewise, the size of every hairpin $\langle m\rangle$ is given by the number of vertices in its corresponding subgraph excluding the entrance vertex (i.e., $|Y(m) \setminus \{m\}|$). Symbol '–' under the hairpin column denotes that no hairpins were found in that graph.

[‡] https://github.com/vgteam/vg/issues/4750