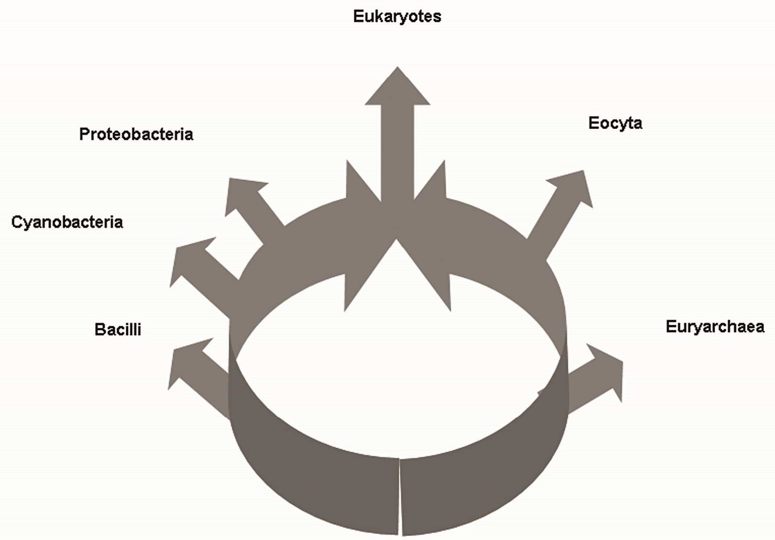
The 'ring of life'

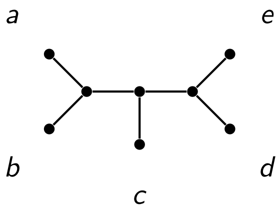
Judith Keijsper and Rudi Pendavingh

Eindhoven

April 24, 2014


Eukaryotes
Proteobacteria
Eocyta
Cyanobacteria
Bacilli
Euryarchaea

Phylogenetic Trees and Quartets

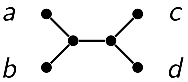


Definition

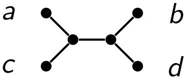
A phylogenetic tree on X is a tree $G = (V, E)$ such that

- X is the set of leaves of G, and
- each $v \in V \setminus X$ has degree 3.

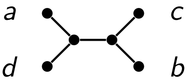
A quartet is a phylogenetic tree on 4 leaves.



$(ab|cd)$

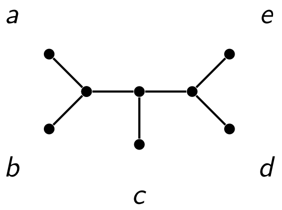


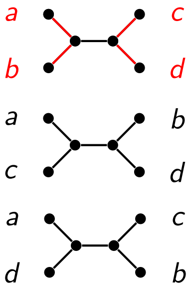
$(ac|bd)$



$(ad|bc)$

Phylogenetic Trees and Quartets





$(ab|cd)$

$(ac|bd)$

$(ad|bc)$

Definition

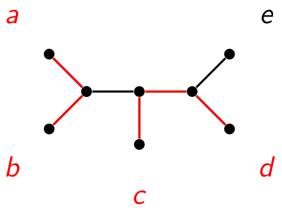
A phylogenetic tree on X is a tree $G = (V, E)$ such that

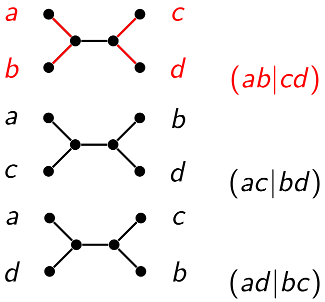
- X is the set of leaves of G, and
- each $v \in V \setminus X$ has degree 3.

A quartet is a phylogenetic tree on 4 leaves.

$(ab|cd)$ denotes the quartet on $\{a, b, c, d\}$ with disjoint paths from a to b and from c to d.

Phylogenetic Trees and Quartets





Definition

A phylogenetic tree on X is a tree $G = (V, E)$ such that

- X is the set of leaves of G, and
- each $v \in V \setminus X$ has degree 3.

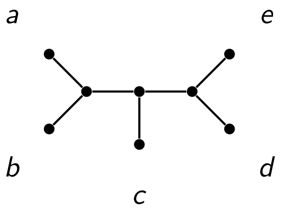
A quartet is a phylogenetic tree on 4 leaves.

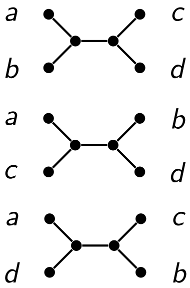
$(ab|cd)$ denotes the quartet on $\{a, b, c, d\}$ with disjoint paths from a to b and from c to d.

Definition

If $a, b, c, d \in X$, then $(ab|cd)$ is consistent with a phylogenetic tree G on X if G contains disjoint paths from a to b and from c to d.

Phylogenetic Trees and Quartets





$(ab|cd)$

$(ac|bd)$

$(ad|bc)$

Definition

A phylogenetic tree on X is a tree $G = (V, E)$ such that

- X is the set of leaves of G, and
- each $v \in V \setminus X$ has degree 3.

A quartet is a phylogenetic tree on 4 leaves.

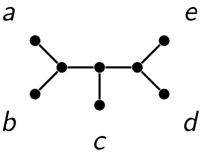
$(ab|cd)$ denotes the quartet on $\{a, b, c, d\}$ with disjoint paths from a to b and from c to d.

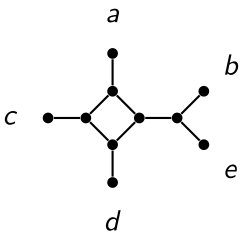
Definition

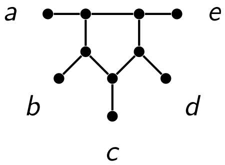
If $a, b, c, d \in X$, then $(ab|cd)$ is consistent with a phylogenetic tree G on X if G contains disjoint paths from a to b and from c to d.

$Q(G)$ denotes the set of quartets consistent with G.

Phylogenetic level-1 Networks





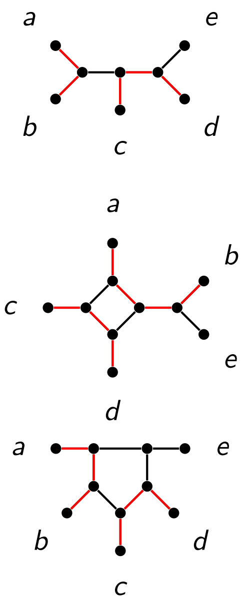


Definition

A level-1 network on X is a connected graph $G = (V, E)$

- in which every two circuits are vertex-disjoint,
- without triangles
- such that X is the set of leaves of G, and
- each $v \in V \setminus X$ has degree 3.

Phylogenetic level-1 Networks



Definition

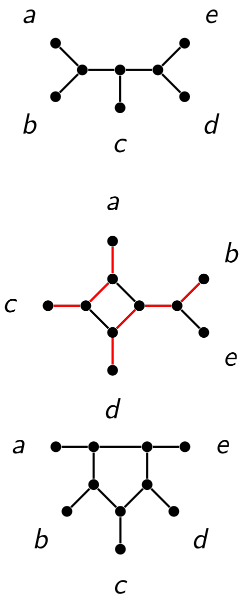
A level-1 network on X is a connected graph $G = (V, E)$

- in which every two circuits are vertex-disjoint,
- without triangles
- such that X is the set of leaves of G, and
- each $v \in V \setminus X$ has degree 3.

Definition

If $a, b, c, d \in X$, then $(ab|cd)$ is *consistent* with a level-1 network G on X if G contains disjoint paths from a to b and from c to d.

Phylogenetic level-1 Networks



Definition

A level-1 network on X is a connected graph $G = (V, E)$

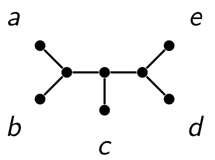
- in which every two circuits are vertex-disjoint,
- without triangles
- such that X is the set of leaves of G, and
- each $v \in V \setminus X$ has degree 3.

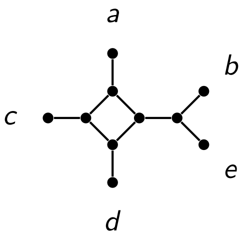
Definition

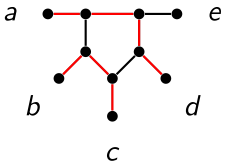
If $a, b, c, d \in X$, then $(ab|cd)$ is *consistent* with a level-1 network G on X if G contains disjoint paths from a to b and from c to d.

$Q(G)$ denotes the set of quartets consistent with G.

Phylogenetic level-1 Networks







Definition

A level-1 network on X is a connected graph $G = (V, E)$

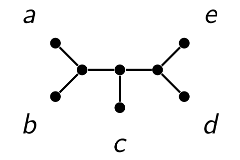
- in which every two circuits are vertex-disjoint,
- without triangles
- such that X is the set of leaves of G, and
- each $v \in V \setminus X$ has degree 3.

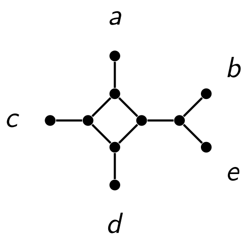
Definition

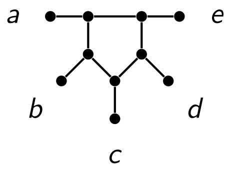
If $a, b, c, d \in X$, then $(ab|cd)$ is *consistent* with a level-1 network G on X if G contains disjoint paths from a to b and from c to d.

$Q(G)$ denotes the set of quartets consistent with G.

Phylogenetic level-1 Networks







Definition

A level-1 network on X is a connected graph $G = (V, E)$

- in which every two circuits are vertex-disjoint,
- without triangles
- such that X is the set of leaves of G, and
- each $v \in V \setminus X$ has degree 3.

Definition

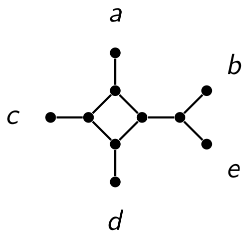
If $a, b, c, d \in X$, then $(ab|cd)$ is *consistent* with a level-1 network G on X if G contains disjoint paths from a to b and from c to d.

$Q(G)$ denotes the set of quartets consistent with G.

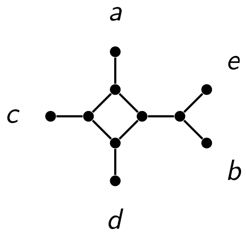
Problem

Given $Q(G)$, reconstruct G in polynomial time.

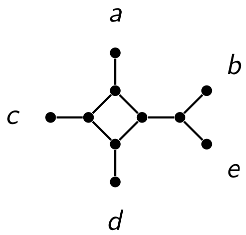
Level-1 Networks and Cyclic orderings



Any level-1 network G on X can be embedded without crossings in a disk with X on the boundary.



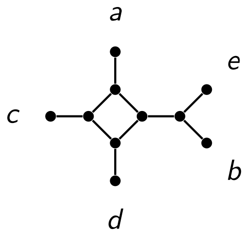
Level-1 Networks and Cyclic orderings



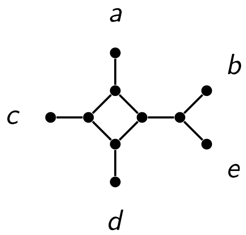
Any level-1 network G on X can be embedded without crossings in a disk with X on the boundary.

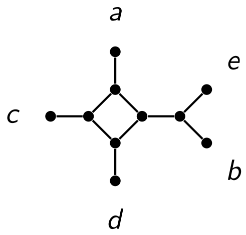
Definition

A cyclic ordering C of X is consistent with G if there is an embedding of G that realizes C (clockwise) on the boundary.



Level-1 Networks and Cyclic orderings





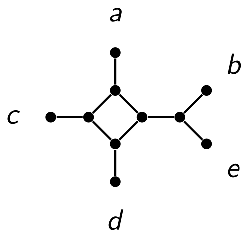
Any level-1 network G on X can be embedded without crossings in a disk with X on the boundary.

Definition

A cyclic ordering C of X is consistent with G if there is an embedding of G that realizes C (clockwise) on the boundary.

$C(G)$ denotes the set of cyclic orderings consistent with G.

Level-1 Networks and Cyclic orderings

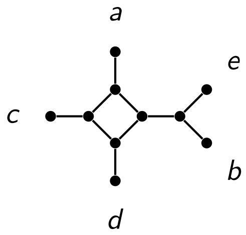


Any level-1 network G on X can be embedded without crossings in a disk with X on the boundary.

Definition

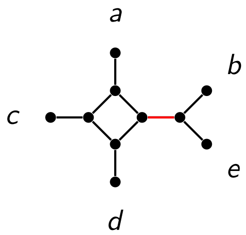
A cyclic ordering C of X is consistent with G if there is an embedding of G that realizes C (clockwise) on the boundary.

$C(G)$ denotes the set of cyclic orderings consistent with G.



In the picture, $C(G) = \{[abedc], [cdeba], [aebdc], [cdbea]\}$.

Level-1 Networks and Cyclic orderings

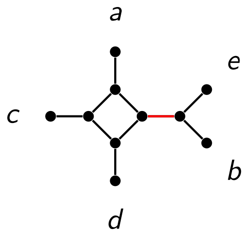


Any level-1 network G on X can be embedded without crossings in a disk with X on the boundary.

Definition

A cyclic ordering C of X is consistent with G if there is an embedding of G that realizes C (clockwise) on the boundary.

$C(G)$ denotes the set of cyclic orderings consistent with G.

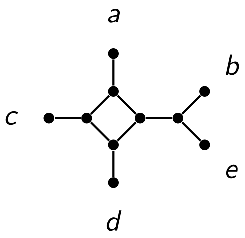


In the picture, $C(G) = \{[abedc], [cdeba], [aebdc], [cdbea]\}$.

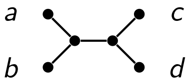
Observation

$|C(G)| = 2^{t+1}$, where t is the number of cut-edges not incident with X.

Level-1 Networks, Cyclic orderings and Quartets



level-1 network G

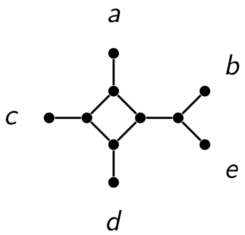


quartet $(ab|cd) \in Q(G)$

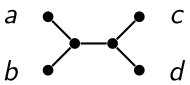
Observation

If $(ab|cd) \in Q(G)$, then a, b, c, d do do not appear in the order $a - c - b - d$ in any cyclic ordering of X consistent with G.

Level-1 Networks, Cyclic orderings and Quartets



level-1 network G



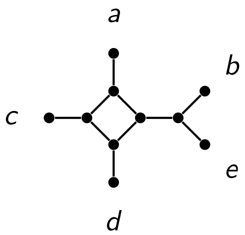
quartet $(ab|cd) \in Q(G)$

Observation

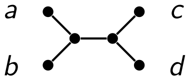
If $(ab|cd) \in Q(G)$, then a, b, c, d do do not appear in the order $a - c - b - d$ in any cyclic ordering of X consistent with G.

A cyclic ordering C is consistent with a quartet $(ab|cd)$ if the pairs (a, b) and (c, d) are not crossing in C.

Level-1 Networks, Cyclic orderings and Quartets



level-1 network G



quartet $(ab|cd) \in Q(G)$

Observation

If $(ab|cd) \in Q(G)$, then a, b, c, d do do not appear in the order $a - c - b - d$ in any cyclic ordering of X consistent with G.

A cyclic ordering C is consistent with a quartet $(ab|cd)$ if the pairs (a, b) and (c, d) are not crossing in C.

Theorem

A cyclic ordering is in $C(G)$ if and only if it is consistent with each quartet in $Q(G)$.

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) \supseteq Q$.

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) \supseteq Q$.

Theorem (Gambette, Berry, Paul 2012)

The following problems are NP-complete:

Given a quartet set Q, decide whether a level-1 network G exists such that $Q(G) \supseteq Q$.

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) \supseteq Q$.

Theorem (Gambette, Berry, Paul 2012)

The following problems are NP-complete:

Given a quartet set Q, decide whether a level-1 network G exists such that $Q(G) \supseteq Q$.

Given a quartet set Q, decide whether a circular ordering consistent with Q exists.

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) \supseteq Q$.

Theorem (Gambette, Berry, Paul 2012)

The following problems are NP-complete:

Given a quartet set Q, decide whether a level-1 network G exists such that $Q(G) \supseteq Q$.

Given a quartet set Q, decide whether a circular ordering consistent with Q exists.

Theorem (GBP 2012)

Given a quartet set Q, it can be decided in polynomial time whether a level-1 network G exists such that

$$Q(G) = Q.$$

If yes, such a network can be constructed in polynomial time.

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Outline of our method:

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Outline of our method:

- from Q, we construct the set $C(G)$ of cyclic orderings consistent with G

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Outline of our method:

- from Q, we construct the set $C(G)$ of cyclic orderings consistent with G
- from $C(G)$, we then recover G

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Outline of our method:

- from Q, we construct the set $C(G)$ of cyclic orderings consistent with G
- from $C(G)$, we then recover G

Main issue: if G is a level-1 network on X, then $C(G)$ may have $2^{|X|-2}$ elements.

Reconstructing a Level-1 Network from Quartets

Problem

Given Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Outline of our method:

- from Q, we construct the set $C(G)$ of cyclic orderings consistent with G
- from $C(G)$, we then recover G

Main issue: if G is a level-1 network on X, then $C(G)$ may have $2^{|X|-2}$ elements.

So, we use an efficient encoding of $C(G)$ to do all this in polynomial time.

Encoding cyclic orderings

A cyclic ordering C of X is uniquely encoded as follows

$$u^C(a,b,c) := \begin{cases} 1 & \text{if } a,b,c \in X \text{ distinct, and clockwise in } C \\ 0 & \text{otherwise} \end{cases}$$

Encoding cyclic orderings

A cyclic ordering C of X is uniquely encoded as follows

$$u^C(a,b,c) := \begin{cases} 1 & \text{if } a,b,c \in X \text{ distinct, and clockwise in } C \\ 0 & \text{otherwise} \end{cases}$$

Theorem (Huntington)

Let $u : X^3 \to GF(2)$. Then u is cyclic, i.e. $u = u^C$ for some cyclic ordering C if and only if

(1) $u(a,a,b) = 0$ *for all* $a, b \in X$ *(distinctness)*

(2) $u(a,b,c) + u(b,c,a) = 0$ *for all* $a,b,c \in X$ *(cyclicity)*

(3) $u(a,b,c) + u(c,b,a) = 1$ *for all* $a,b,c \in X$ *(asymmetry)*

(4) $u(a,b,c)u(a,c,d) = u(a,b,d)u(b,c,d)$ *for all* $a,b,c,d \in X$ *(transitivity)*

Encoding cyclic orderings

A cyclic ordering C of X is uniquely encoded as follows

$$u^C(a,b,c) := \begin{cases} 1 & \text{if } a,b,c \in X \text{ distinct, and clockwise in } C \\ 0 & \text{otherwise} \end{cases}$$

Theorem (Huntington)

Let $u : X^3 \to GF(2)$. Then u is cyclic, i.e. $u = u^C$ for some cyclic ordering C if and only if

(1) $u(a,a,b) = 0$ *for all* $a,b \in X$ *(distinctness)*

(2) $u(a,b,c) + u(b,c,a) = 0$ *for all* $a,b,c \in X$ *(cyclicity)*

(3) $u(a,b,c) + u(c,b,a) = 1$ *for all* $a,b,c \in X$ *(asymmetry)*

(4) $u(a,b,c)u(a,c,d) = u(a,b,d)u(b,c,d)$ *for all* $a,b,c,d \in X$ *(transitivity)*

Another linear equation that is valid for cyclic vectors u:

(5) $u(a,b,c) + u(a,b,d) + u(a,c,d) + u(b,c,d) = 0$

Encoding cyclic orderings

A cyclic ordering C of X is uniquely encoded as follows

$$u^C(a,b,c) := \begin{cases} 1 & \text{if } a,b,c \in X \text{ distinct, and clockwise in } C \\ 0 & \text{otherwise} \end{cases}$$

Theorem (Huntington)

Let $u : X^3 \to GF(2)$. Then u is cyclic, i.e. $u = u^C$ for some cyclic ordering C if and only if

(1) $u(a,a,b) = 0$ *for all* $a, b \in X$ *(distinctness)*

(2) $u(a,b,c) + u(b,c,a) = 0$ *for all* $a,b,c \in X$ *(cyclicity)*

(3) $u(a,b,c) + u(c,b,a) = 1$ *for all* $a,b,c \in X$ *(asymmetry)*

(4) $u(a,b,c)u(a,c,d) = u(a,b,d)u(b,c,d)$ *for all* $a,b,c,d \in X$ *(transitivity)*

Another linear equation that is valid for cyclic vectors u:

(5) $u(a,b,c) + u(a,b,d) + u(a,c,d) + u(b,c,d) = 0$

$$\mathcal{U}^X := \{u \in GF(2)^{X^3} : u \text{ satisfies (1),(2),(3), and (5) } \}$$

An affine space over GF(2)

Observation

A cyclic ordering C is consistent with a quartet $(ab|cd)$ if and only if $u^C(a, b, c) = u^C(a, b, d)$

An affine space over GF(2)

Observation

A cyclic ordering C is consistent with a quartet $(ab|cd)$ if and only if $u^C(a,b,c) = u^C(a,b,d)$

For a set of quartets on X, define

$$U(Q) := \{u \in \mathcal{U}^X : u(a,b,c) = u(a,b,d) \text{ for all } (ab|cd) \in Q\}$$

An affine space over GF(2)

Observation

A cyclic ordering C is consistent with a quartet $(ab|cd)$ if and only if $u^C(a,b,c) = u^C(a,b,d)$

For a set of quartets on X, define

$$U(Q) := \{u \in \mathcal{U}^X : u(a,b,c) = u(a,b,d) \text{ for all } (ab|cd) \in Q\}$$

If G is a level-1 network on X, we write

$$U(G) := \{u^C : C \in C(G)\}$$

An affine space over GF(2)

Observation

A cyclic ordering C is consistent with a quartet $(ab|cd)$ if and only if $u^C(a,b,c) = u^C(a,b,d)$

For a set of quartets on X, define

$$U(Q) := \{u \in \mathcal{U}^X : u(a,b,c) = u(a,b,d) \text{ for all } (ab|cd) \in Q\}$$

If G is a level-1 network on X, we write

$$U(G) := \{u^C : C \in C(G)\}$$

Theorem

$U(G) = U(Q(G))$

An affine space over GF(2)

Observation

A cyclic ordering C is consistent with a quartet $(ab|cd)$ if and only if $u^C(a,b,c) = u^C(a,b,d)$

For a set of quartets on X, define

$$U(Q) := \{u \in \mathcal{U}^X : u(a,b,c) = u(a,b,d) \text{ for all } (ab|cd) \in Q\}$$

If G is a level-1 network on X, we write

$$U(G) := \{u^C : C \in C(G)\}$$

Theorem

$U(G) = U(Q(G))$

So in particular, $U(G)$ is an affine subspace of $GF(2)^{X^3}$.

Our algorithm

Problem

Given a quartet set Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Our algorithm

Problem

Given a quartet set Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Algorithm:

Our algorithm

Problem

Given a quartet set Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Algorithm:

- from Q, compute $U(Q) := \{u \in \mathcal{U}^X : u(a,b,c) = u(a,b,d) \text{ for all } (ab|cd) \in Q\}$

Our algorithm

Problem

Given a quartet set Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Algorithm:

- from Q, compute $U(Q) := \{u \in \mathcal{U}^X : u(a,b,c) = u(a,b,d) \text{ for all } (ab|cd) \in Q\}$
- from $U(G) = U(Q)$, we then recover G.

Our algorithm

Problem

Given a quartet set Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Algorithm:

- from Q, compute $U(Q) := \{u \in \mathcal{U}^X : u(a,b,c) = u(a,b,d) \text{ for all } (ab|cd) \in Q\}$
- from $U(G) = U(Q)$, we then recover G.

Theorem

Our algorithm takes $O(|X|^6)$ time.

Our algorithm

Problem

Given a quartet set Q, reconstruct a level-1 network G such that $Q(G) = Q$, in polynomial time.

Algorithm:

- from Q, compute $U(Q) := \{u \in \mathcal{U}^X : u(a,b,c) = u(a,b,d) \text{ for all } (ab|cd) \in Q\}$
- from $U(G) = U(Q)$, we then recover G.

Theorem

Our algorithm takes $O(|X|^6)$ time.

Proof: The linear algebra over $GF(2)$ involves $|Q| = O(|X|^4)$ sparse equations in $\dim(\mathcal{U}^X) = \binom{|X|-1}{2}$ variables.

The dense case

A set of quartets Q on X is dense if for every $a, b, c, d \in X$, at least one of the quartets $(ab|cd)$, $(ac|bd)$ or $(ad|bc)$ is in Q.

The dense case

A set of quartets Q on X is dense if for every $a, b, c, d \in X$, at least one of the quartets $(ab|cd)$, $(ac|bd)$ or $(ad|bc)$ is in Q.

Observation

If Q is dense, then $U(Q)$ contains only cyclic vectors.

The dense case

A set of quartets Q on X is dense if for every $a, b, c, d \in X$, at least one of the quartets $(ab|cd)$, $(ac|bd)$ or $(ad|bc)$ is in Q.

Observation

If Q is dense, then $U(Q)$ contains only cyclic vectors.

Proof: the quartet equation

$$u(a, b, c) = u(a, b, d)$$

together with equation (5) implies transitivity (equation (4)) for the quadruple a, b, c, d.

The dense case

A set of quartets Q on X is dense if for every $a, b, c, d \in X$, at least one of the quartets $(ab|cd)$, $(ac|bd)$ or $(ad|bc)$ is in Q.

Observation

If Q is dense, then $U(Q)$ contains only cyclic vectors.

Proof: the quartet equation

$$u(a, b, c) = u(a, b, d)$$

together with equation (5) implies transitivity (equation (4)) for the quadruple a, b, c, d.

Problems (GBP 2012)

Given a dense quartet set Q, reconstruct a level-1 network G such that $Q(G) \supseteq Q$.

The dense case

A set of quartets Q on X is dense if for every $a, b, c, d \in X$, at least one of the quartets $(ab|cd)$, $(ac|bd)$ or $(ad|bc)$ is in Q.

Observation

If Q is dense, then $U(Q)$ contains only cyclic vectors.

Proof: the quartet equation

$$u(a, b, c) = u(a, b, d)$$

together with equation (5) implies transitivity (equation (4)) for the quadruple a, b, c, d.

Problems (GBP 2012)

Given a dense quartet set Q, reconstruct a level-1 network G such that $Q(G) \supseteq Q$.

Given a dense quartet set Q on X, find a cyclic ordering consistent with Q.

The dense case

A set of quartets Q on X is dense if for every $a, b, c, d \in X$, at least one of the quartets $(ab|cd)$, $(ac|bd)$ or $(ad|bc)$ is in Q.

Observation

If Q is dense, then $U(Q)$ contains only cyclic vectors.

Proof: the quartet equation

$$u(a, b, c) = u(a, b, d)$$

together with equation (5) implies transitivity (equation (4)) for the quadruple a, b, c, d.

Problems (GBP 2012)

Given a dense quartet set Q, reconstruct a level-1 network G such that $Q(G) \supseteq Q$.

Given a dense quartet set Q on X, find a cyclic ordering consistent with Q.

Theorem

These problems can be solved in polynomial time.

Recovering the most tree-like G from $U(G)$

If G is a level-1 network on X, then $S \subseteq X$ is a split of G if

$$S = X \cap U$$

where $|\delta_G(U)| = 1$. We write $\Sigma(G)$ for the set of splits of G.
For a set $S \subseteq X$, we define the vector $v : X^3 \to GF(2)$ by

$$v^S(x,y,z) := \begin{cases} 1 & x,y,z \text{ distinct, and } |\{x,y,z\} \cap S| \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

Recovering the most tree-like G from $U(G)$

If G is a level-1 network on X, then $S \subseteq X$ is a split of G if

$$S = X \cap U$$

where $|\delta_G(U)| = 1$. We write $\Sigma(G)$ for the set of splits of G.
For a set $S \subseteq X$, we define the vector $v : X^3 \to GF(2)$ by

$$v^S(x,y,z) := \begin{cases} 1 & x,y,z \text{ distinct, and } |\{x,y,z\} \cap S| \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

Theorem

$U(G) = u^C + lin.hull\{v^S : S \text{ a split of } G\}$, where C is some cyclic ordering of G.

Recovering the most tree-like G from $U(G)$

If G is a level-1 network on X, then $S \subseteq X$ is a split of G if

$$S = X \cap U$$

where $|\delta_G(U)| = 1$. We write $\Sigma(G)$ for the set of splits of G.
For a set $S \subseteq X$, we define the vector $v : X^3 \to GF(2)$ by

$$v^S(x, y, z) := \begin{cases} 1 & x, y, z \text{ distinct, and } |\{x, y, z\} \cap S| \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

Theorem

$U(G) = u^C + lin.hull\{v^S : S \text{ a split of } G\}$, where C is some cyclic ordering of G.

So if we know $U(G)$, then we can

Recovering the most tree-like G from $U(G)$

If G is a level-1 network on X, then $S \subseteq X$ is a split of G if

$$S = X \cap U$$

where $|\delta_G(U)| = 1$. We write $\Sigma(G)$ for the set of splits of G.

For a set $S \subseteq X$, we define the vector $v : X^3 \rightarrow GF(2)$ by

$$v^S(x, y, z) := \begin{cases} 1 & x, y, z \text{ distinct, and } |\{x, y, z\} \cap S| \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

Theorem

$U(G) = u^C + lin.hull\{v^S : S$ a split of $G\}$, where C is some cyclic ordering of G.

So if we know $U(G)$, then we can

- recover some $C \in C(G)$ from a $u \in U(G)$

Recovering the most tree-like G from $U(G)$

If G is a level-1 network on X, then $S \subseteq X$ is a split of G if

$$S = X \cap U$$

where $|\delta_G(U)| = 1$. We write $\Sigma(G)$ for the set of splits of G.
For a set $S \subseteq X$, we define the vector $v : X^3 \to GF(2)$ by

$$v^S(x,y,z) := \begin{cases} 1 & x,y,z \text{ distinct, and } |\{x,y,z\} \cap S| \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

Theorem

$U(G) = u^C + lin.hull\{v^S : S$ a split of $G\}$, where C is some cyclic ordering of G.

So if we know $U(G)$, then we can

- recover some $C \in C(G)$ from a $u \in U(G)$
- recover $\Sigma(G) = \{S \subseteq X : S \text{ is consecutive in } C, \text{ and such that } u^C + v^S \in U(G)\}$

Recovering the most tree-like G from $U(G)$

If G is a level-1 network on X, then $S \subseteq X$ is a split of G if

$$S = X \cap U$$

where $|\delta_G(U)| = 1$. We write $\Sigma(G)$ for the set of splits of G.

For a set $S \subseteq X$, we define the vector $v : X^3 \to GF(2)$ by

$$v^S(x,y,z) := \begin{cases} 1 & x,y,z \text{ distinct, and } |\{x,y,z\} \cap S| \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

Theorem

$U(G) = u^C + lin.hull\{v^S : S$ a split of $G\}$, where C is some cyclic ordering of G.

So if we know $U(G)$, then we can

- recover some $C \in C(G)$ from a $u \in U(G)$
- recover $\Sigma(G) = \{S \subseteq X : S$ is consecutive in C, and such that $u^C + v^S \in U(G)\}$
- recover G from C and Σ

Computing $U(G)$ from fewer quartets

Lemma

Let G be a level-1 network on X where $|X| = 5$. For an $x \in X$, let

$$Q_x := \{(ab|cd) \in Q(G) : x \in \{a, b, c, d\}\}$$

Then $U(G) = U(Q_x)$.

Computing $U(G)$ from fewer quartets

Lemma

Let G be a level-1 network on X where $|X| = 5$. For an $x \in X$, let

$$Q_x := \{(ab|cd) \in Q(G) : x \in \{a, b, c, d\}\}$$

Then $U(G) = U(Q_x)$.

So to recover G, we may not need the full set of quartets.

Computing $U(G)$ from fewer quartets

Lemma

Let G be a level-1 network on X where $|X| = 5$. For an $x \in X$, let

$$Q_x := \{(ab|cd) \in Q(G) : x \in \{a, b, c, d\}\}$$

Then $U(G) = U(Q_x)$.

So to recover G, we may not need the full set of quartets.

Theorem

Let G be a level-1 network on X. For an $x \in X$, let

$$Q_x := \{(ab|cd) \in Q(G) : x \in \{a, b, c, d\}\}$$

Then $U(G) = U(Q_x)$.

The general case

A variant of our algorithm solves the following problem

Problem

Given: *A (general) quartet set* Q *on* X, *and some* $x \in X$.
Find: *a level-1 network* G *with the maximum number of splits such that* $Q(G) \supseteq Q$
or *a certificate of inconsistency* $Q' \subseteq Q$, *where* $|Q'| \leq \binom{|X|-1}{2}$
or *a 4-tuple* $Z \subseteq X$ *with* $x \in Z$ *that is a witness of the fact that* $U(Q)$ *contains non-cyclic vectors.*

The general case

A variant of our algorithm solves the following problem

Problem

Given: *A (general) quartet set* Q *on* X, *and some* $x \in X$.
Find: *a level-1 network* G *with the maximum number of splits such that* $Q(G) \supseteq Q$
or *a certificate of inconsistency* $Q' \subseteq Q$, *where* $|Q'| \leq \binom{|X|-1}{2}$
or *a 4-tuple* $Z \subseteq X$ *with* $x \in Z$ *that is a witness of the fact that* $U(Q)$ *contains non-cyclic vectors.*

Interactive use:

Keep adding new quartets on the 4-tuples Z returned by the algorithm until either a level-1 network G or an inconsistent subset of quartets Q' is obtained.