Clustering Systems of Phylogenetic Networks

Marc Hellmuth

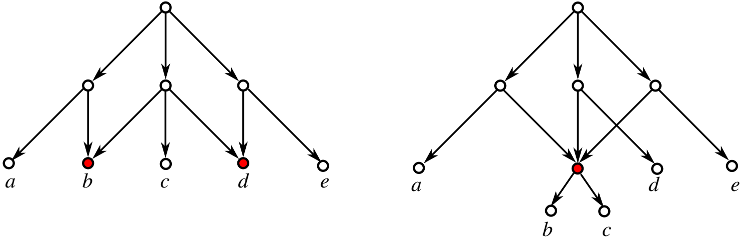
Department of Mathematics
Faculty of Science
Stockholm University

TBI Winterseminar, 2024

Basics

All networks $N = (V, E)$ considered here are

1. DAGs with a single root
2. phylogenetic (= no indegree 1 and outdegree 1 vertices)

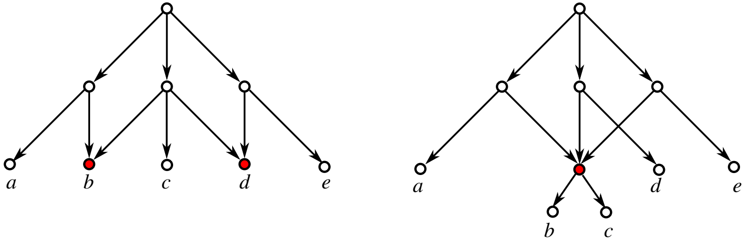


hybrid-vertex = ● / leaf = outdegree 0

Basics

All networks $N = (V, E)$ considered here are

1. DAGs with a single root
2. phylogenetic (= no indegree 1 and outdegree 1 vertices)



hybrid-vertex = ● / leaf = outdegree 0

(In)comparable vertices

$u \preceq_N v$ if v is an ancestor of u, i.e., there is a directed path from v to u.

u and v are $\preceq_N$-comparable if $u \preceq_N v$ or $v \preceq_N u$

Otherwise, u and v are $\preceq_N$-incomparable

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves “below” v.

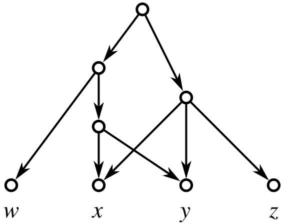
Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves “below” v.



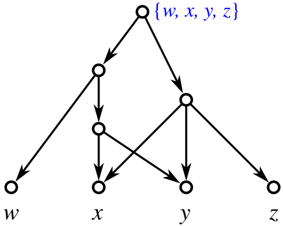
Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.



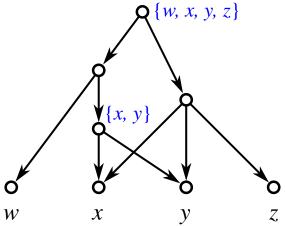
Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.



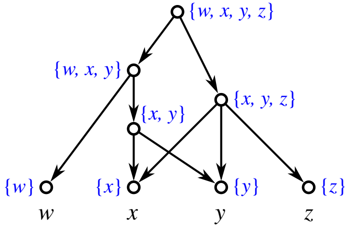
Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.



$$\mathscr{C}_N = \Big\{\{w,x,y,z\},\{w,x,y\},\{x,y,z\},\{x,y\},\{w\},\{x\},\{y\},\{z\}\Big\}$$

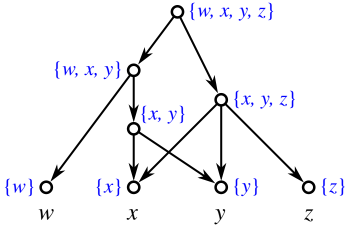
Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.



$$\mathscr{C}_N = \Big\{\{w,x,y,z\},\{w,x,y\},\{x,y,z\},\{x,y\},\{w\},\{x\},\{y\},\{z\}\Big\}$$

How much information about N is contained in $\mathscr{C}_N$?

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

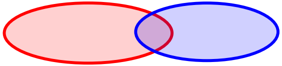
$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.

Folklore (Trees):

- $C, C' \in \mathscr{C}$ **overlap**, if $C \cap C' \notin \{C, C', \emptyset\}$.

  A clustering system is a **hierarchy** if it does not contain pairwise overlapping sets.



OVERLAP

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V,E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

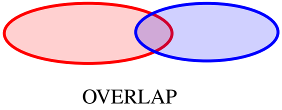
where $C(v)$ denotes the set of leaves "below" v.

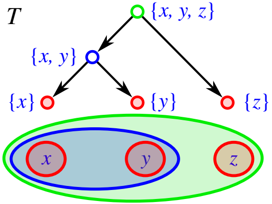
Folklore (Trees):

- $C, C' \in \mathscr{C}$ **overlap**, if $C \cap C' \notin \{C, C', \emptyset\}$.

  A clustering system is a **hierarchy** if it does not contain pairwise overlapping sets.

- **There is a 1-to-1 correspondence between rooted phylogenetic trees and hierarchies.**
  [keyword: Hasse-diagram]





Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

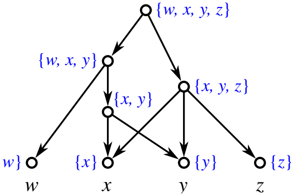
For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.

Observations about networks:

- $\mathscr{C}_N$ is usually *not* a hierarchy for general networks



Here: $\mathscr{C} = \{\{w,x,y,z\}, \{w,x,y\}, \{x,y,z\}, \{y,z\}, \{w\}, \{x\}, \{y\}, \{z\}\}$

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

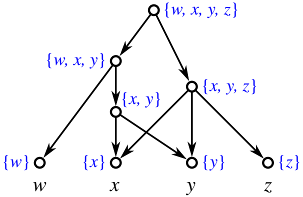
For all $N = (V, E)$, there is a *unique* clustering system

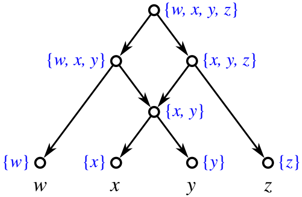
$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.

Observations about networks:

- $\mathscr{C}_N$ is usually *not* a hierarchy for general networks





Here: $\mathscr{C} = \{\{w,x,y,z\}, \{w,x,y\}, \{x,y,z\}, \{y,z\}, \{w\}, \{x\}, \{y\}, \{z\}\}$

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

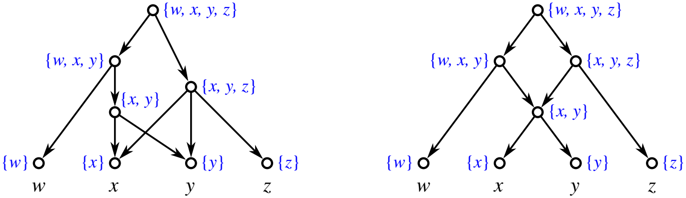
For all $N = (V,E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.

Observations about networks:

- $\mathscr{C}_N$ is usually *not* a hierarchy for general networks
- *No* 1-to-1 correspondence between networks and clustering systems.



Here: $\mathscr{C} = \{\{w,x,y,z\}, \{w,x,y\}, \{x,y,z\}, \{y,z\}, \{w\}, \{x\}, \{y\}, \{z\}\}$

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves “below” v.

Central Questions: Given $\mathscr{C}$.

- When is there a network N of a given type such that $\mathscr{C} = \mathscr{C}_N$?

  type = tree, level-k, binary, galled tree, tree-child, tree-based, regular, normal, . . .

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.

Central Questions: Given $\mathscr{C}$.

- When is there a network N of a given type such that $\mathscr{C} = \mathscr{C}_N$?

  type = tree, level-k, binary, galled tree, tree-child, tree-based, regular, normal, . . .
- Which type of networks are uniquely determined by $\mathscr{C}_N$?

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V,E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves “below” v.

Central Questions: Given $\mathscr{C}$.

- When is there a network N of a given type such that $\mathscr{C} = \mathscr{C}_N$?
  type = tree, level-k, binary, galled tree, tree-child, tree-based, regular, normal, . . .
- Which type of networks are uniquely determined by $\mathscr{C}_N$?

Clustering systems

A **clustering system** on X is a set $\mathscr{C} \subseteq 2^X \setminus \{\emptyset\}$ such that $X \in \mathscr{C}$ and $\{x\} \in \mathscr{C}$ for all $x \in X$.

For all $N = (V, E)$, there is a *unique* clustering system

$$\mathscr{C}_N := \{C(v) \mid v \in V\}$$

where $C(v)$ denotes the set of leaves "below" v.

Central Questions: Given $\mathscr{C}$.

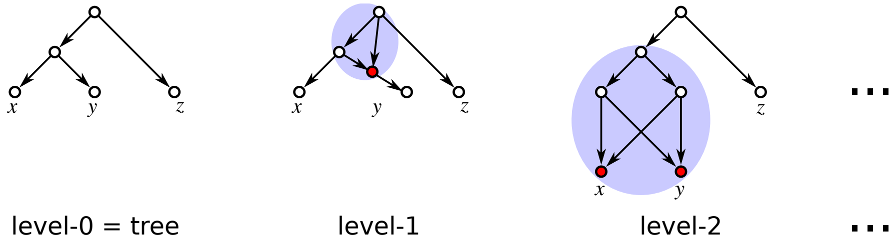
- When is there a network N of a given type such that $\mathscr{C} = \mathscr{C}_N$?
  type = tree, level-k, binary, galled tree, tree-child, tree-based, regular, normal, ...
- Which type of networks are uniquely determined by $\mathscr{C}_N$?

Let us try to answer some of the questions for level-1 networks.

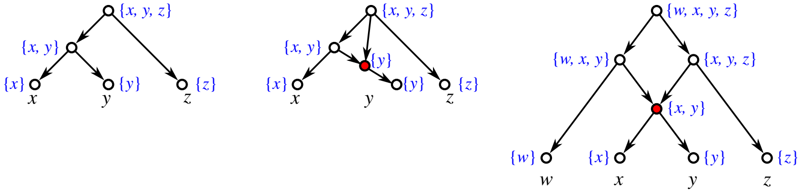
Level-k networks

A **block B** in a network is a maximal biconnected subgraph.

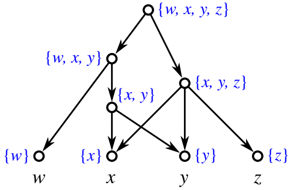
A network N is **level-k** if each block B in N contains $\leq k$ hybrid-vertices
(distinct from root ρ_B of B).



Level-1 Networks



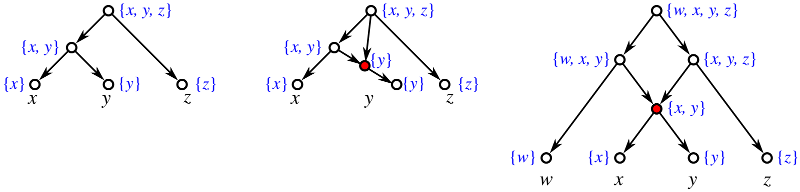
Level-1



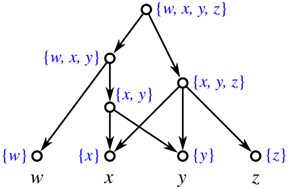
not Level-1

In trees T: $\quad u \preceq_T v \iff C(u) \subseteq C(v)$

Level-1 Networks



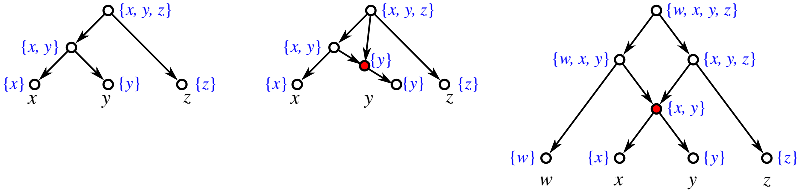
Level-1



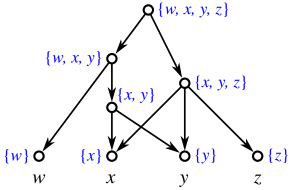
not Level-1

In trees T: $\quad u \preceq_T v \iff C(u) \subseteq C(v)$

Level-1 Networks



Level-1

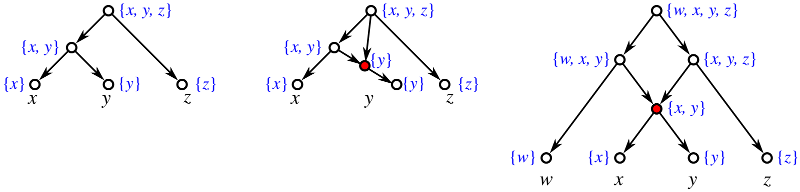


not Level-1

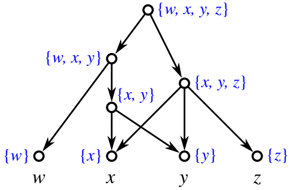
In trees T: $u \preceq_T v \iff C(u) \subseteq C(v)$

In networks N: $u \preceq_N v \implies C(u) \subseteq C(v)$ *(converse not true in general)*

Level-1 Networks



Level-1



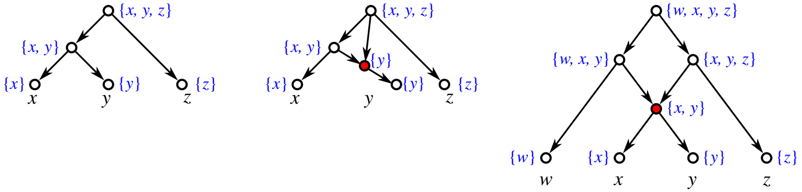
not Level-1

In trees T: $u \preceq_T v \iff C(u) \subseteq C(v)$

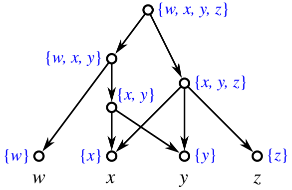
In networks N: $u \preceq_N v \implies C(u) \subseteq C(v)$ *(converse not true in general)*

Hence, if $C(u)$ and $C(v)$ overlap or are disjoint, then u and v are $\preceq_N$-incomparable in N

Level-1 Networks



Level-1



not Level-1

In trees T: $u \preceq_T v \iff C(u) \subseteq C(v)$

In networks N: $u \preceq_N v \implies C(u) \subseteq C(v)$ *(converse not true in general)*

Hence, if $C(u)$ and $C(v)$ overlap or are disjoint, then u and v are $\preceq_N$-incomparable in N

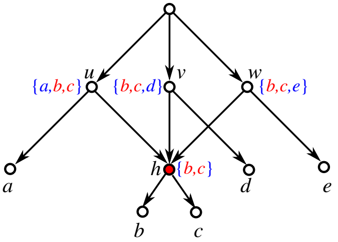
Lemma (H., Schaller, Stadler, 2023)

In level-1 networks N: u and v are $\preceq_N$-incomparable in N $\iff$ $C(u) \cap C(v) \in \{\emptyset, C(h_B)\}$ where $h_B \neq u, v$ is the unique hybrid in block B that contains u and v.

Property (L)

If $C(u)$ overlaps with $C(v)$, then $C(u) \cap C(v) = C(h_B)$ with B being the block containing u and v.

[this property cannot be observed when just looking at $\mathscr{C}_N$!]

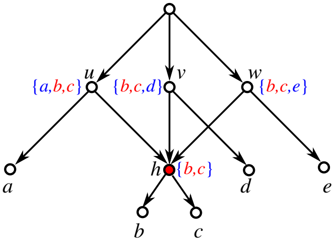


Property (L)

If $C(u)$ overlaps with $C(v)$, then $C(u) \cap C(v) = C(h_B)$ with B being the block containing u and v.

[this property cannot be observed when just looking at $\mathscr{C}_N$!]

A clustering system $\mathscr{C}$ satisfies **property (L)** if
$C \cap C_1 = C \cap C_2$ for all $C, C_1, C_2 \in \mathscr{C}$ where C overlaps both C_1 and C_2.



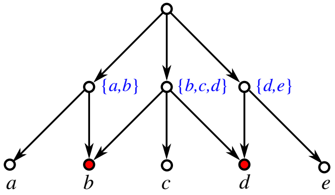
property (L)

Property (L)

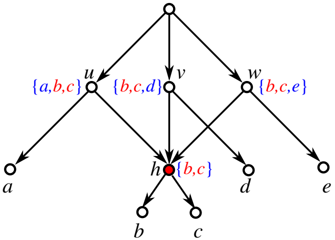
If $C(u)$ overlaps with $C(v)$, then $C(u) \cap C(v) = C(h_B)$ with B being the block containing u and v.

[this property cannot be observed when just looking at $\mathscr{C}_N$!]

A clustering system $\mathscr{C}$ satisfies **property (L)** if
$C \cap C_1 = C \cap C_2$ for all $C, C_1, C_2 \in \mathscr{C}$ where C overlaps both C_1 and C_2.



no property (L)



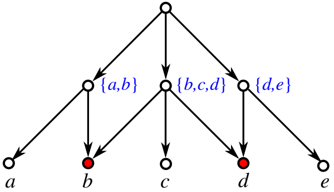
property (L)

Property (L)

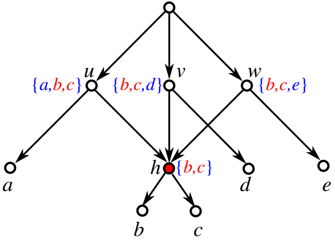
If $C(u)$ overlaps with $C(v)$, then $C(u) \cap C(v) = C(h_B)$ with B being the block containing u and v.

[this property cannot be observed when just looking at $\mathscr{C}_N$!]

A clustering system $\mathscr{C}$ satisfies **property (L)** if
$C \cap C_1 = C \cap C_2$ for all $C, C_1, C_2 \in \mathscr{C}$ where C overlaps both C_1 and C_2.



no property (L)



property (L)

Lemma (H., Schaller, Stadler, 2023)

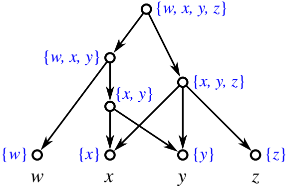
For all level-1 networks N, the set $\mathscr{C}_N$ satisfies property (L).

Closed clustering systems

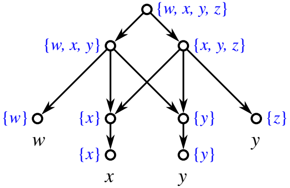
A clustering system $\mathscr{C}$ is **closed** if $A \cap B \in \mathscr{C}$ for all $A, B \in \mathscr{C}$ with $A \cap B \neq \emptyset$.

Closed clustering systems

A clustering system $\mathscr{C}$ is **closed** if $A \cap B \in \mathscr{C}$ for all $A, B \in \mathscr{C}$ with $A \cap B \neq \emptyset$.



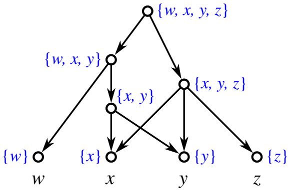
closed



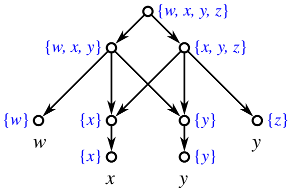
not closed

Closed clustering systems

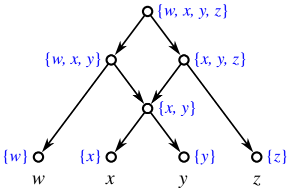
A clustering system $\mathscr{C}$ is **closed** if $A \cap B \in \mathscr{C}$ for all $A, B \in \mathscr{C}$ with $A \cap B \neq \emptyset$.



closed



not closed



level-1 = closed

Lemma (H., Schaller, Stadler, 2023)

For level-1 networks N, the set $\mathscr{C}_N$ is always closed.

Characterization for level-1 Networks

A clustering system $\mathscr{C}$ ...

... is **closed** if $A \cap B \in \mathscr{C}$ for all $A, B \in \mathscr{C}$ with $A \cap B \neq \emptyset$.

... satisfies **property (L)** if $C \cap C_1 = C \cap C_2$ for all $C, C_1, C_2 \in \mathscr{C}$ where C overlaps both C_1 and C_2.

Characterization for level-1 Networks

A clustering system $\mathscr{C}$ ...

... is **closed** if $A \cap B \in \mathscr{C}$ for all $A, B \in \mathscr{C}$ with $A \cap B \neq \emptyset$.

... satisfies **property (L)** if $C \cap C_1 = C \cap C_2$ for all $C, C_1, C_2 \in \mathscr{C}$ where C overlaps both C_1 and C_2.

Theorem (H., Schaller, Stadler, 2023)

For $\mathscr{C}$, there is a level-1 network N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is closed and satisfies (L).

Characterization for level-1 Networks

A clustering system $\mathscr{C}$ ...

... is **closed** if $A \cap B \in \mathscr{C}$ for all $A, B \in \mathscr{C}$ with $A \cap B \neq \emptyset$.

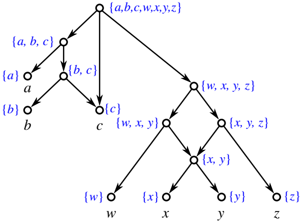
... satisfies **property (L)** if $C \cap C_1 = C \cap C_2$ for all $C, C_1, C_2 \in \mathscr{C}$ where C overlaps both C_1 and C_2.

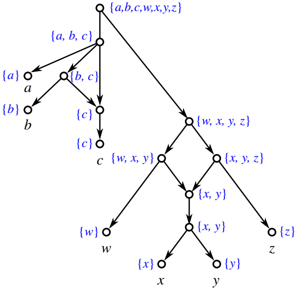
Theorem (H., Schaller, Stadler, 2023)

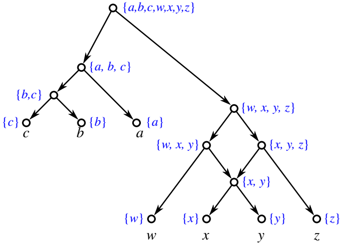
For $\mathscr{C}$, there is a level-1 network N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is closed and satisfies (L).

Recognition of such $\mathscr{C}$ and reconstruction of such a level-1 network can be done in polynomial time.

Uniqueness Results



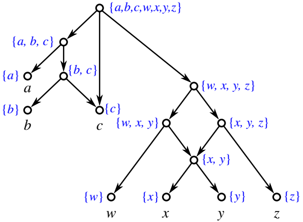


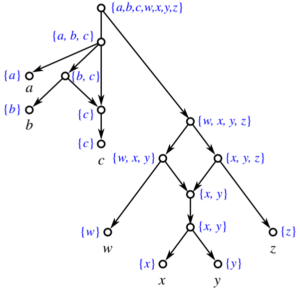


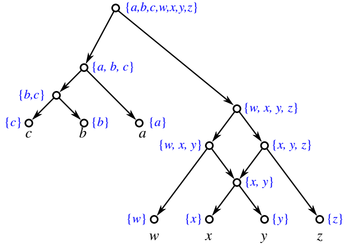
Hasse diagram

In general, several different level-1 networks may represent the same $\mathscr{C}$

Uniqueness Results







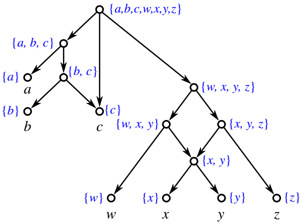
Hasse diagram

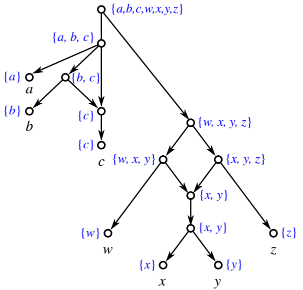
In general, several different level-1 networks may represent the same $\mathscr{C}$

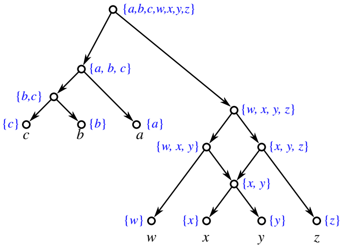
Proposition (H., Schaller, Stadler, 2023)

Let $\mathscr{C}$ be a closed clustering system that satisfies (L). Then, the Hasse-diagram of $\mathscr{C}$ is the unique least-resolved level-1 network N with $\mathscr{C}_N = \mathscr{C}$.

Uniqueness Results







Hasse diagram

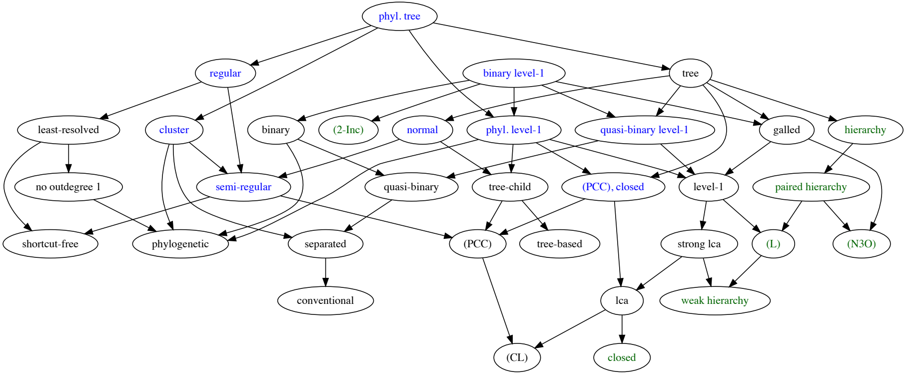
In general, several different level-1 networks may represent the same $\mathscr{C}$

Proposition (H., Schaller, Stadler, 2023)

Let $\mathscr{C}$ be a closed clustering system that satisfies (L). Then, the Hasse-diagram of $\mathscr{C}$ is the unique least-resolved level-1 network N with $\mathscr{C}_N = \mathscr{C}$.

Every level-1 network N is a refinement (=adding shortcuts + expand vertices) of the Hasse-diagram of $\mathscr{C}_N$.

Other types of networks . . .



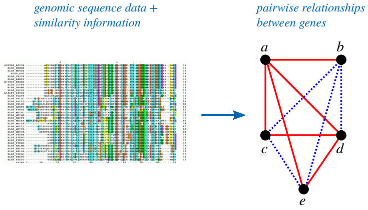
The latter results and plenty of other characterizations for dozens of other network types can be found in
*Hellmuth, Stadler, Schaller, **Clustering Systems of Phylogenetic Networks**, Theory in Biosciences (142), 301-358, 2023*

The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology

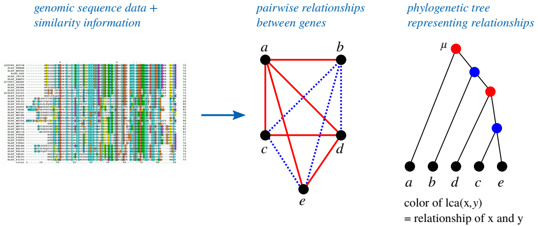
The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology



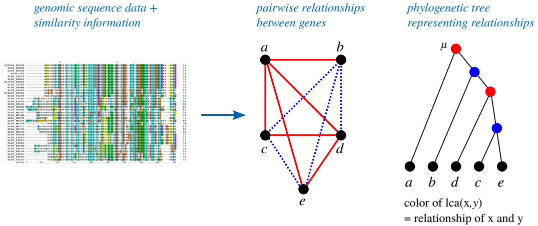
The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology



The initial motivation to investigate the clustering systems of networks:

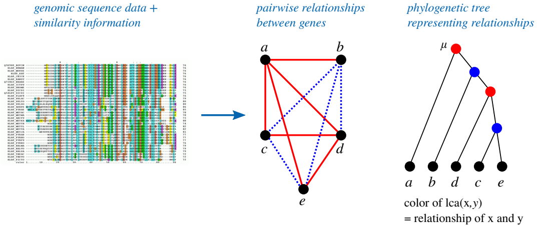
- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology



Given a relationship R that can be represented by trees T

The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology

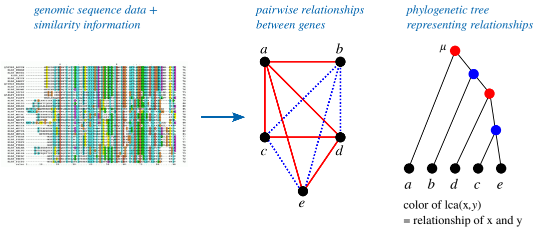


Given a relationship R that can be represented by trees T

- Clustering system of T is determined by subgraphs (modules) of R

The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology

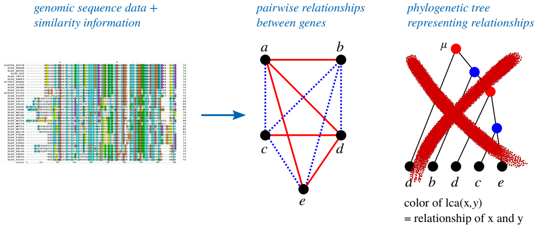


Given a relationship R that can be represented by trees T

- Clustering system of T is determined by subgraphs (modules) of R
- Colors of pairwise lca's determine relationship

The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology

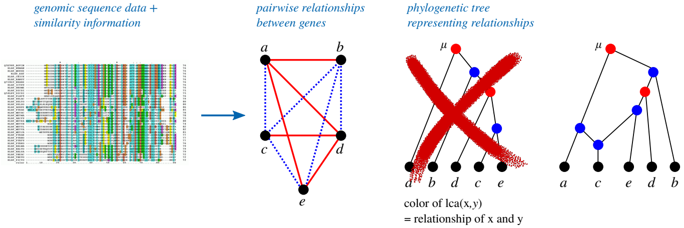


Given a relationship R that can be represented by trees T

- Clustering system of T is determined by subgraphs (modules) of R
- Colors of pairwise lca's determine relationship
- Noise in the data or NON-tree-like evolution $\Longrightarrow$ cannot expect trees !

The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology

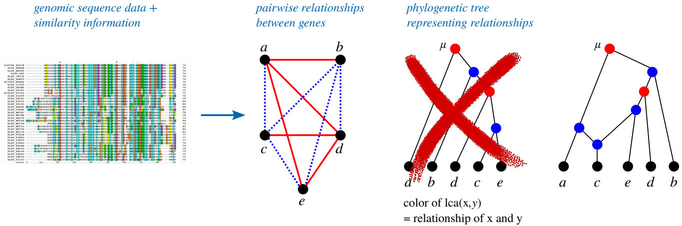


Given a relationship R that can be represented by trees T

- Clustering system of T is determined by subgraphs (modules) of R
- Colors of pairwise lca's determine relationship
- Noise in the data or NON-tree-like evolution $\Longrightarrow$ cannot expect trees !

The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology



Given a relationship R that can be represented by trees T

- Clustering system of T is determined by subgraphs (modules) of R

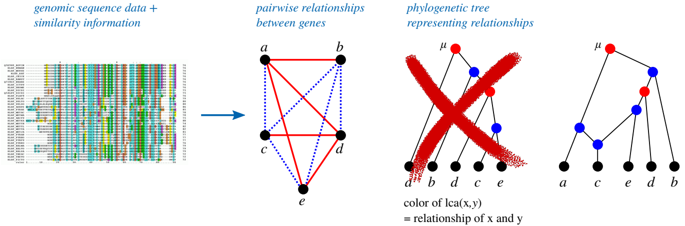
  How are subgraphs in R related to clusters in N ?

- Colors of pairwise lca's determine relationship

- Noise in the data or NON-tree-like evolution $\Longrightarrow$ cannot expect trees !

The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology



Given a relationship R that can be represented by trees T

- Clustering system of T is determined by subgraphs (modules) of R

  How are subgraphs in R related to clusters in N ?

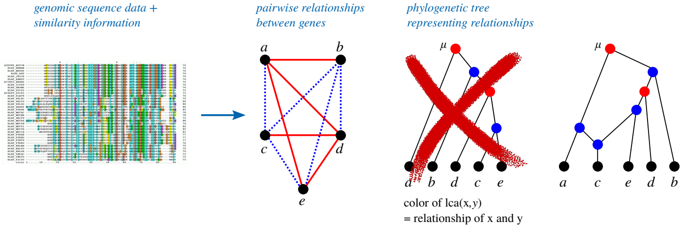
- Colors of pairwise lca's determine relationship

  How are networks characterized with pairwise-lca-properties?

- Noise in the data or NON-tree-like evolution $\Longrightarrow$ cannot expect trees !

The initial motivation to investigate the clustering systems of networks:

- lack of results in literature
- we wanted to understand in more detail inference of horizontal gene transfer and orthology



Given a relationship R that can be represented by trees T

- Clustering system of T is determined by subgraphs (modules) of R

  How are subgraphs in R related to clusters in N ?

- Colors of pairwise lca's determine relationship

  How are networks characterized with pairwise-lca-properties?

- Noise in the data or NON-tree-like evolution $\Longrightarrow$ cannot expect trees !

A first starting point is provided by

Theorem (Shanavas, Changat, H., Stadler, 2024)

For $\mathscr{C}$, there is a network N with $\mathscr{C} = \mathscr{C}_N$ and pairwise lca*-property*
$\iff$ *$\mathscr{C}_N$ is pre-binary, i.e., there is a unique incl.-min. cluster $C \in \mathscr{C}$ such that $\{x,y\} \subseteq C$ for all $x,y \in L(N)$*

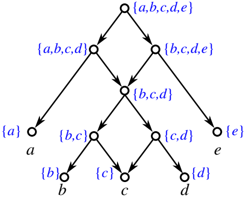
Shanavas, Changat, Hellmuth, Stadler ***Unique Least Common Ancestors and Clusters in Directed Acyclic Graphs****, LNCS vol 14508, 2024*

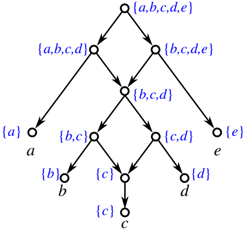
- David Schaller (Biontech, GER)
- Peter F. Stadler (Uni Leipzig, GER)
- Ameera Vaheeda Shanavas (Uni Kerala, IND)
- Manoj Changa (Uni Kerala, IND)
- Anna Lindeberg (Uni Stockholm, SWE)

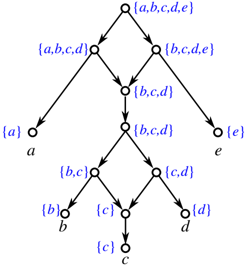
- David Schaller (Biontech, GER)
- Peter F. Stadler (Uni Leipzig, GER)
- Ameera Vaheeda Shanavas (Uni Kerala, IND)
- Manoj Changa (Uni Kerala, IND)
- Anna Lindeberg (Uni Stockholm, SWE)

Thanks!

Uniqueness Results







We obtain uniqueness for mild restrictions!

Theorem

Let $\mathscr{C}$ be a closed clustering system that satisfies (L). Then, there is a unique shortcut-free level-1 network N with $\mathscr{C}_N = \mathscr{C}$ that satisfies precisely one condition:

- *N contains no vertex v with outdegree 1 (=Hasse diagram)*
- *every leaf in N has indegree 1 and all vertices v with outdegree 1 are adjacent to leaves.*
- *every hybrid in N has outdegree 1 (thus all leaves have indegree 1).*

Further results

Galled tree = level-1 network where all non-trivial blocks correspond to “undirected cycles”

Further results

Galled tree = level-1 network where all non-trivial blocks correspond to “undirected cycles”

Theorem

*There is a galled tree N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is **closed** and satisfies **Property (L)** and does not contain three distinct pairwise overlapping clusters.*

Further results

Galled tree = level-1 network where all non-trivial blocks correspond to “undirected cycles”

Theorem

*There is a galled tree N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is **closed** and satisfies **Property (L)** and does not contain three distinct pairwise overlapping clusters.*

Binary network = network where all non-hybrids v is either a leaf or has $\operatorname{outdeg}(v) = 2$, and every hybrid v satisfies $\operatorname{indeg}(v) = 2$ and $\operatorname{outdeg}(\mathrm{v}) = 1$.

Further results

Galled tree = level-1 network where all non-trivial blocks correspond to “undirected cycles”

Theorem

*There is a galled tree N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is **closed** and satisfies **Property (L)** and does not contain three distinct pairwise overlapping clusters.*

Binary network = network where all non-hybrids v is either a leaf or has $\mathrm{outdeg}(v) = 2$, and every hybrid v satisfies $\mathrm{indeg}(v) = 2$ and $\mathrm{outdeg}(\mathrm{v}) = 1$.

Theorem

*There is a binary level-1 network N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is **closed** and satisfies **Property (L)** and for all clusters $C \in \mathscr{C}$, there are at most two inclusion-maximal clusters $A, B \in \mathscr{C}$ with $A, B \subsetneq C$ and at most two inclusion-minimal clusters $A, B \in \mathscr{C}$ with $C \subsetneq A, B$.*

Further results

Galled tree = level-1 network where all non-trivial blocks correspond to “undirected cycles”

Theorem

*There is a galled tree N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is **closed** and satisfies **Property (L)** and does not contain three distinct pairwise overlapping clusters.*

Binary network = network where all non-hybrids v is either a leaf or has $\operatorname{outdeg}(v) = 2$, and every hybrid v satisfies $\operatorname{indeg}(v) = 2$ and $\operatorname{outdeg}(\mathrm{v}) = 1$.

Theorem

*There is a binary level-1 network N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is **closed** and satisfies **Property (L)** and for all clusters $C \in \mathscr{C}$, there are at most two inclusion-maximal clusters $A, B \in \mathscr{C}$ with $A, B \subsetneq C$ and at most two inclusion-minimal clusters $A, B \in \mathscr{C}$ with $C \subsetneq A, B$.*

A clustering system $\mathscr{C}$ is compatible w.r.t. N if $\mathscr{C} \subseteq \mathscr{C}_N$.

Further results

Galled tree = level-1 network where all non-trivial blocks correspond to “undirected cycles”

Theorem

*There is a galled tree N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is **closed** and satisfies **Property (L)** and does not contain three distinct pairwise overlapping clusters.*

Binary network = network where all non-hybrids v is either a leaf or has $\mathrm{outdeg}(v) = 2$, and every hybrid v satisfies $\mathrm{indeg}(v) = 2$ and $\mathrm{outdeg}(\mathrm{v}) = 1$.

Theorem

*There is a binary level-1 network N with $\mathscr{C}_N = \mathscr{C} \iff \mathscr{C}$ is **closed** and satisfies **Property (L)** and for all clusters $C \in \mathscr{C}$, there are at most two inclusion-maximal clusters $A, B \in \mathscr{C}$ with $A, B \subsetneq C$ and at most two inclusion-minimal clusters $A, B \in \mathscr{C}$ with $C \subsetneq A, B$.*

A clustering system $\mathscr{C}$ is compatible w.r.t. N if $\mathscr{C} \subseteq \mathscr{C}_N$.

Theorem

*$\mathscr{C}$ is compatible w.r.t. a level-1 network $N \iff \mathscr{C}$ is satisfies **Property (L)**.*

In this case, compute $A \cap B$ for all overlapping $A, B \in \mathscr{C}$ and add it to $\mathscr{C}$ if their intersection is not present. (can be done in polynomial time)

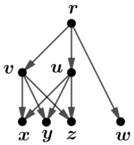


Fig. 2. Consider the DAG G with leaf set $X = L(G)$ where $\mathscr{C}_G = \{\{x\}, \{y\}, \{z\}, \{w\}, \{x, y, z\}, X\}$. Here, $\mathscr{C}_G$ satisfies **(KS)** and **(KC)** for $k = 2$. By definition, $\mathscr{C}_G$ is thus pre-binary. However, G is not a pairwise lca-network since $\operatorname{lca}(x, y)$, $\operatorname{lca}(x, z)$, and $\operatorname{lca}(y, z)$ are not defined. Moreover, $\mathscr{C}_G$ also satisfies **(KC)** for $k = 3$ but G is not a *3*-lca-network since $\operatorname{lca}(x, y, z)$ is not defined.

pre-binary clustering system but not pairwise lca-property