40 années d'algorithmique de graphes
40 Years of Graphs and Algorithms
11/10/2018 - Paris

From graph classes to phylogenetic networks

Philippe Gambette

UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Gaspard Monge Laboratoire d'informatique

Outline

- Discovering graph classes with Michel
- An introduction to phylogenetic networks
- Classes of phylogenetic networks
- The Tree Containment Problem
- Counting phylogenetic networks

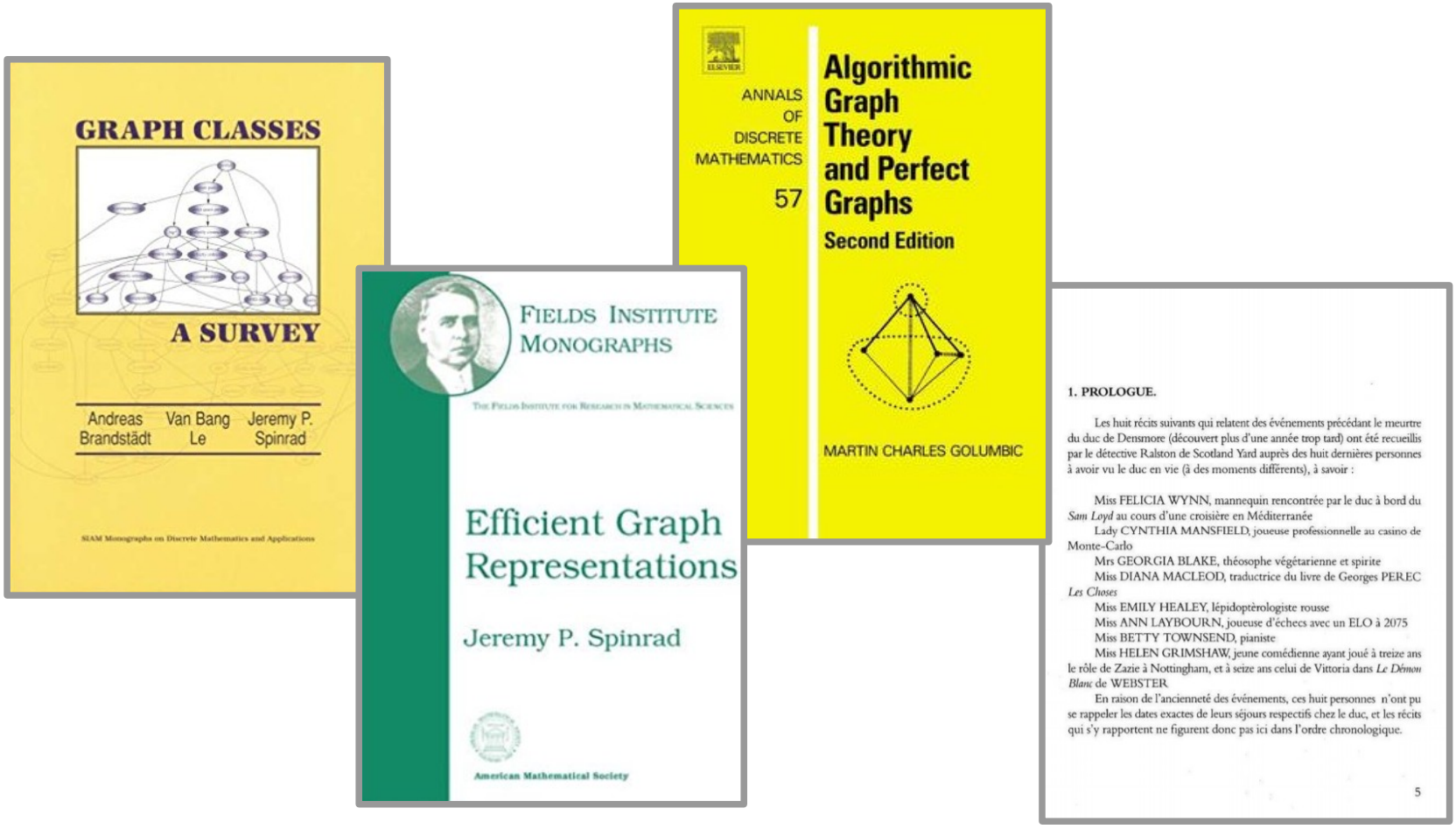
Outline

- Discovering graph classes with Michel
- An introduction to phylogenetic networks
- Classes of phylogenetic networks
- The Tree Containment Problem
- Counting phylogenetic networks

Discovering graph classes with Michel

Master thesis supervised by Michel Habib about 2-interval graphs

→ Reading about graph classes, learning about graph decompositions



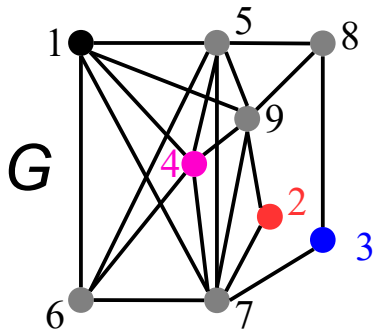
Discovering graph classes with Michel

Master thesis supervised by Michel Habib about 2-interval graphs

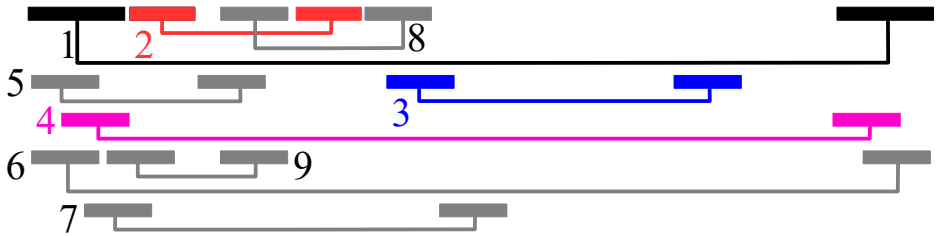
→ Reading about graph classes, learning about graph decompositions

→ Obtaining results on balanced 2-interval graphs and number of arc-annotated sequences

A balanced 2-interval graph *G*:



A balanced 2-interval realization of *G*:



a vertex ↔ balanced 2-intervals
= union of 2 intervals of same length

an edge ↔ non empty intersection of the two 2-intervals

Discovering graph classes with Michel

Master thesis supervised by Michel Habib about 2-interval graphs

→ Reading about graph classes, learning about graph decompositions

→ Obtaining results on balanced 2-interval graphs and number of arc-annotated sequences

→ Understanding what is important about the obtained results

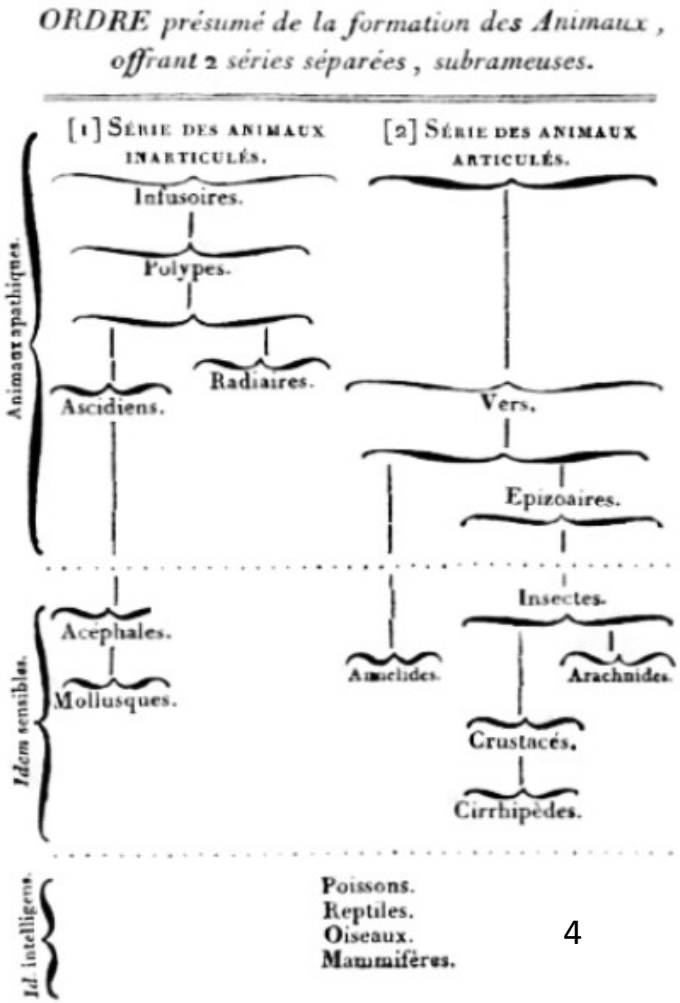
Outline

- Discovering graph classes with Michel
- An introduction to phylogenetic networks
- Classes of phylogenetic networks
- The Tree Containment Problem
- Counting phylogenetic networks

Phylogenetic trees

Phylogenetic tree of a set of species:

- **Classify them depending on common characters** **classification**
- Describe their evolution

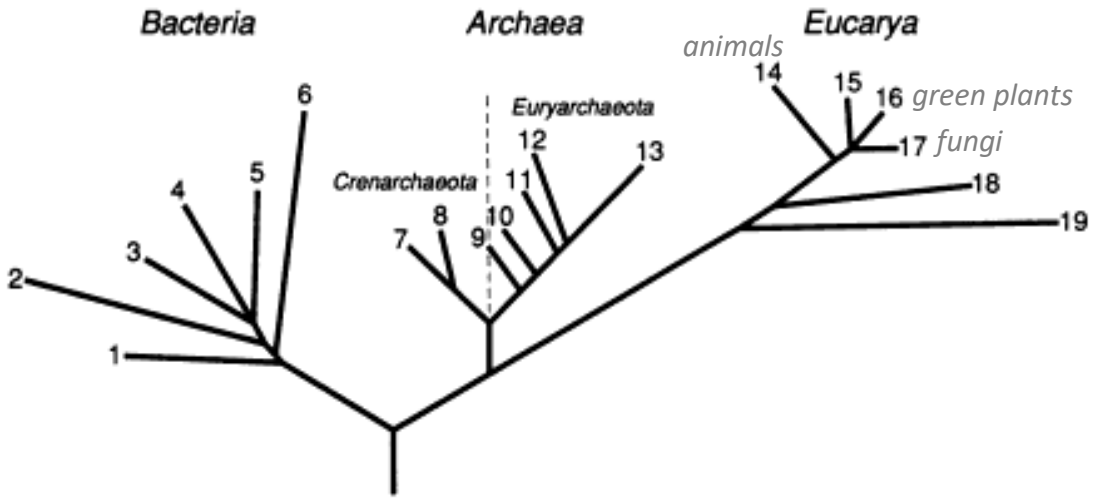


Lamarck : Histoire naturelle des animaux sans vertèbres (1815)

Phylogenetic trees

Phylogenetic tree of a set of species:

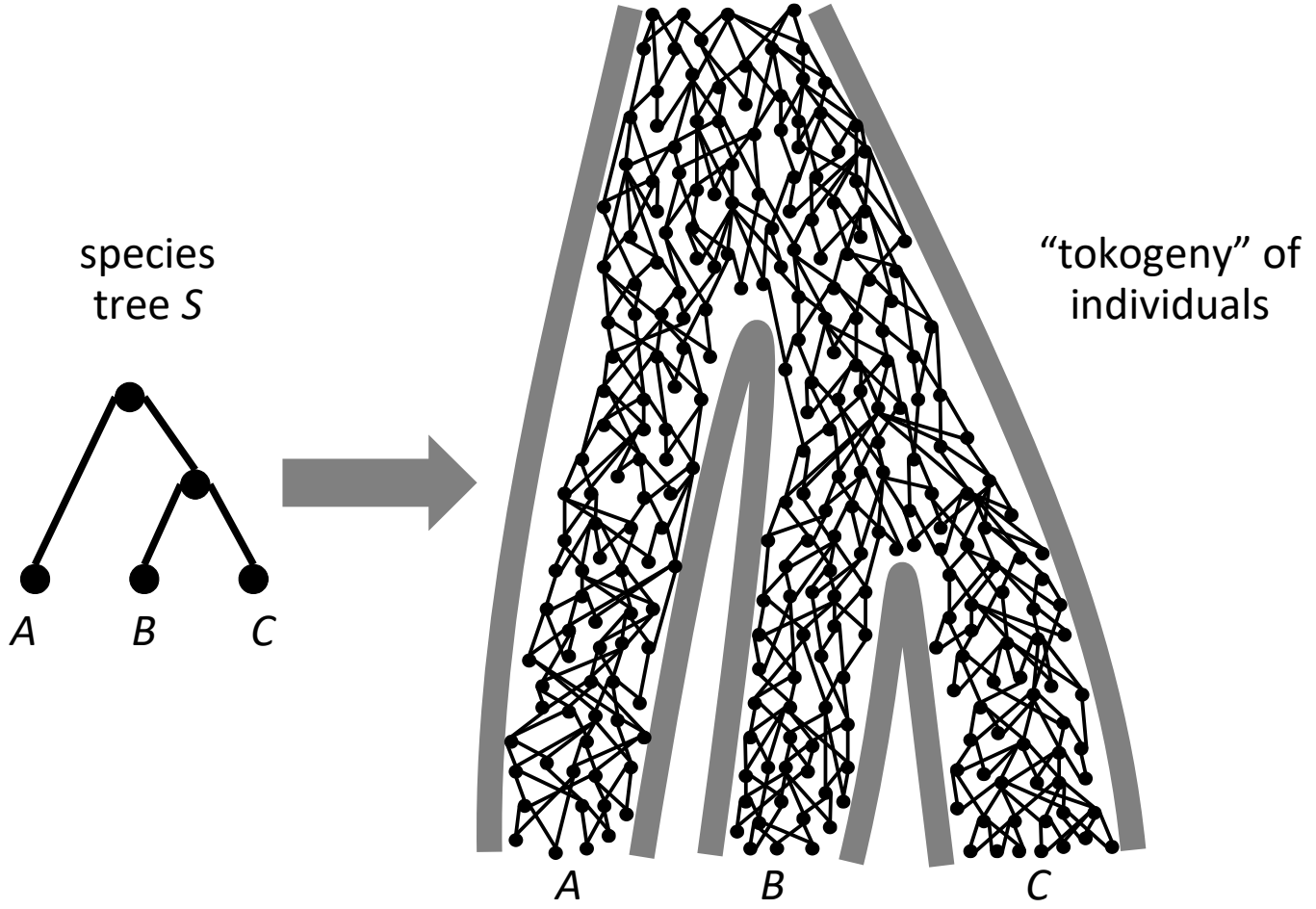
- Classify them depending on common characters
- **Describe their evolution** **modelization**



Woese, Kandler, Wheelis : Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, Proceedings of the National Academy of Sciences, 87(12), 4576–4579 (1990)

Phylogenetic trees... and networks

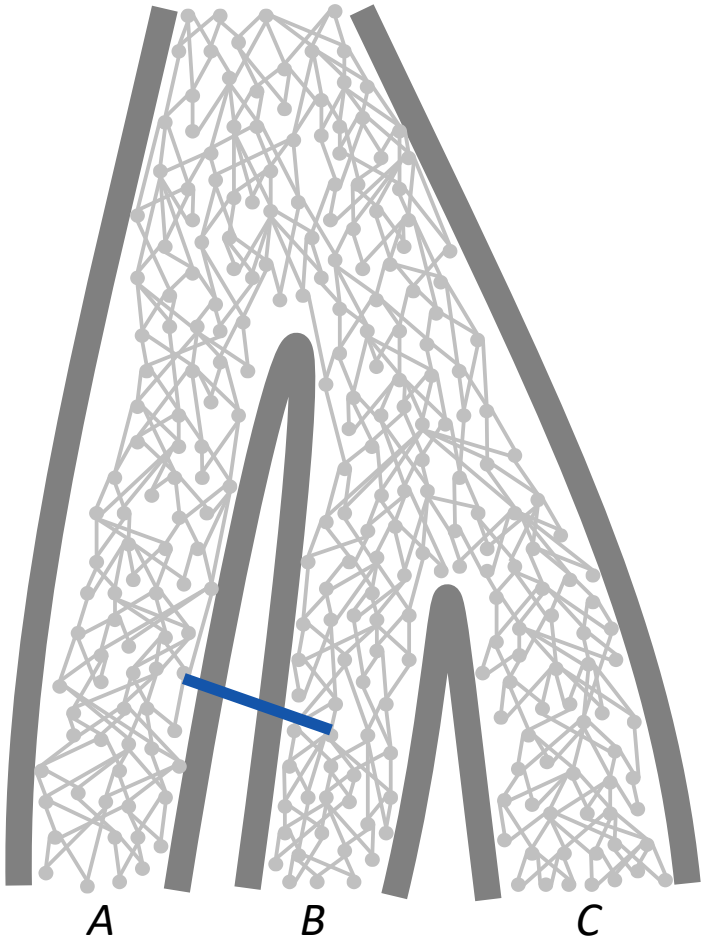
Phylogenetic tree of a set of species



Genetic material transfers

Transfers of genetic material between coexisting species:

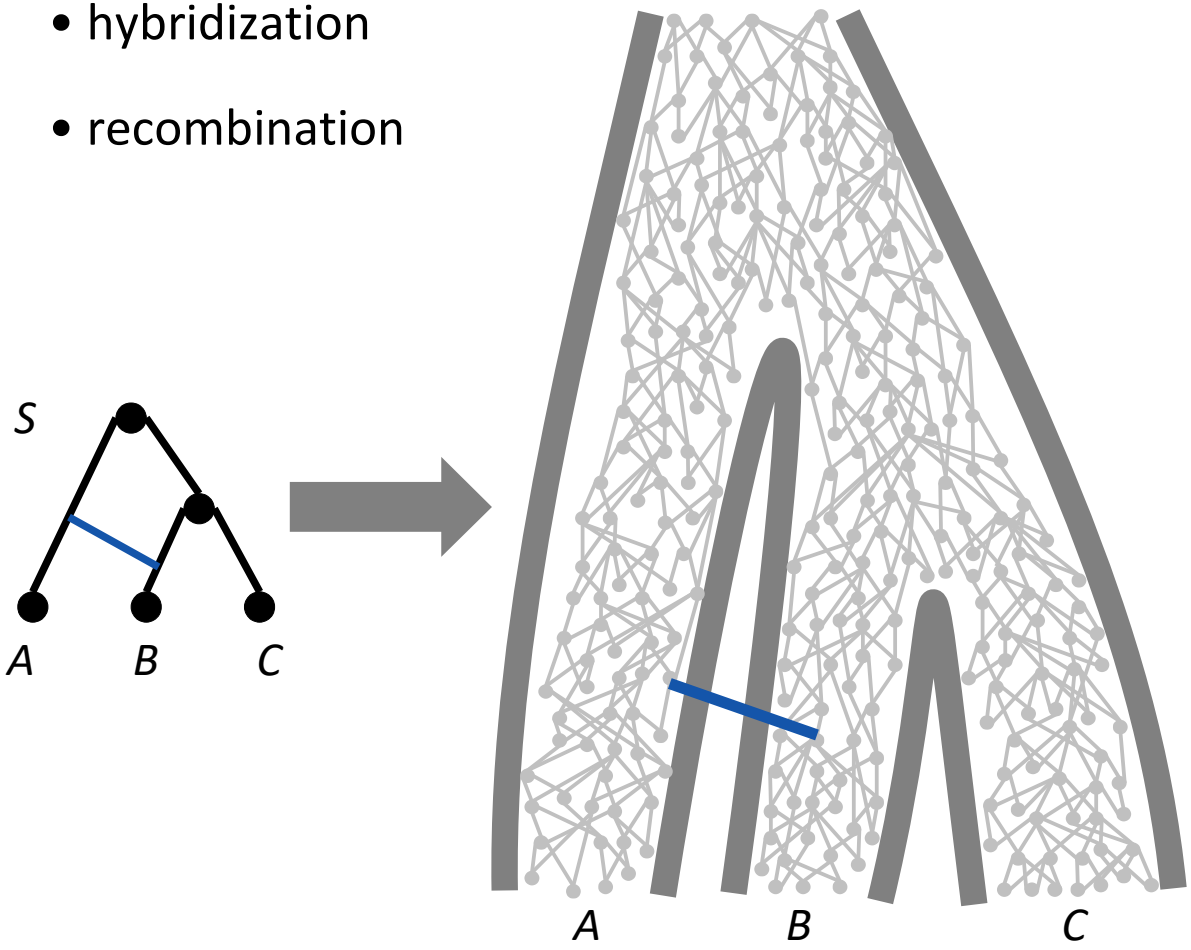
- lateral gene transfer
- hybridization
- recombination



Genetic material transfers

Transfers of genetic material between coexisting species:

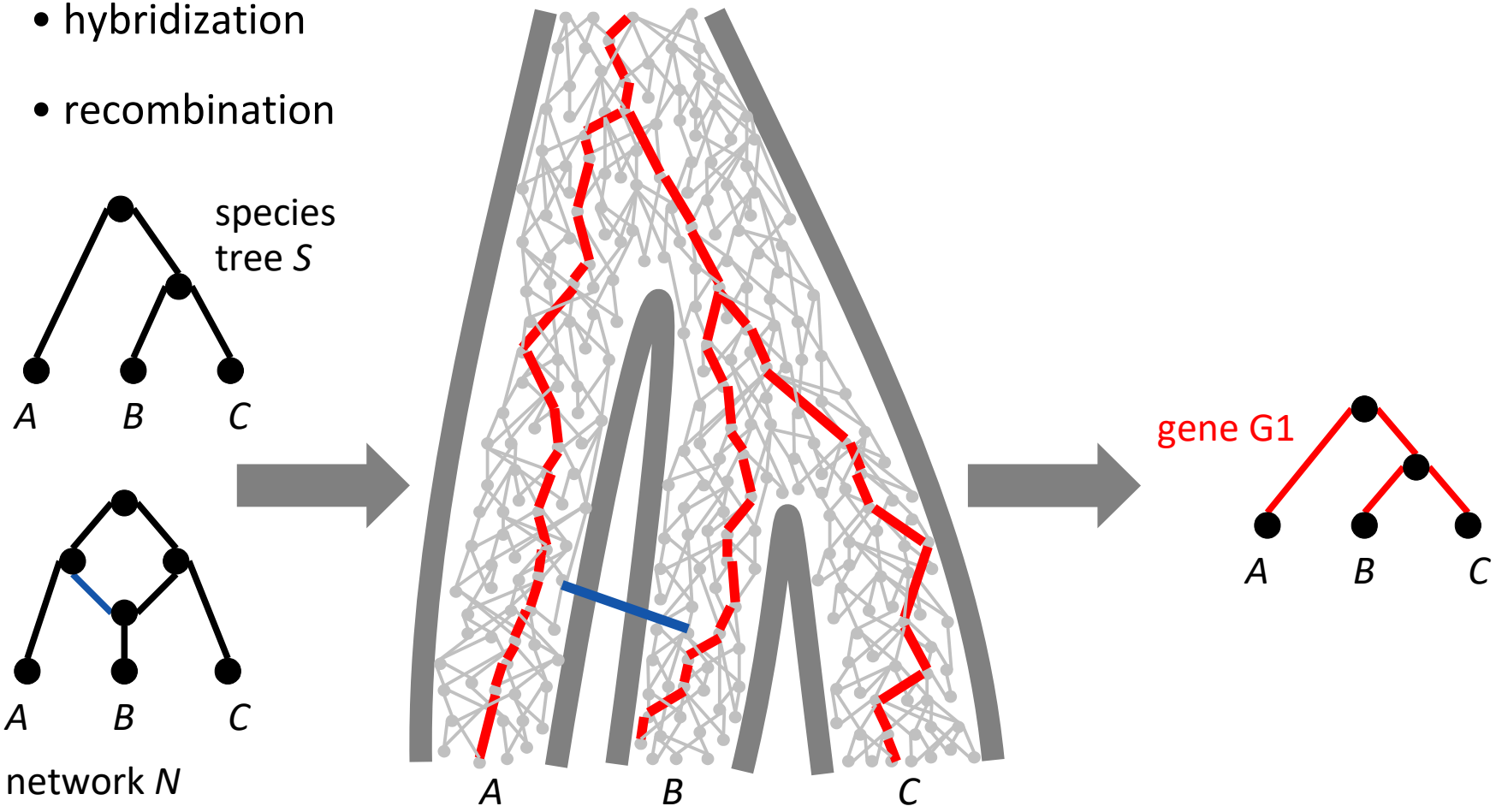
- lateral gene transfer
- hybridization
- recombination



Genetic material transfers

Transfers of genetic material between coexisting species:

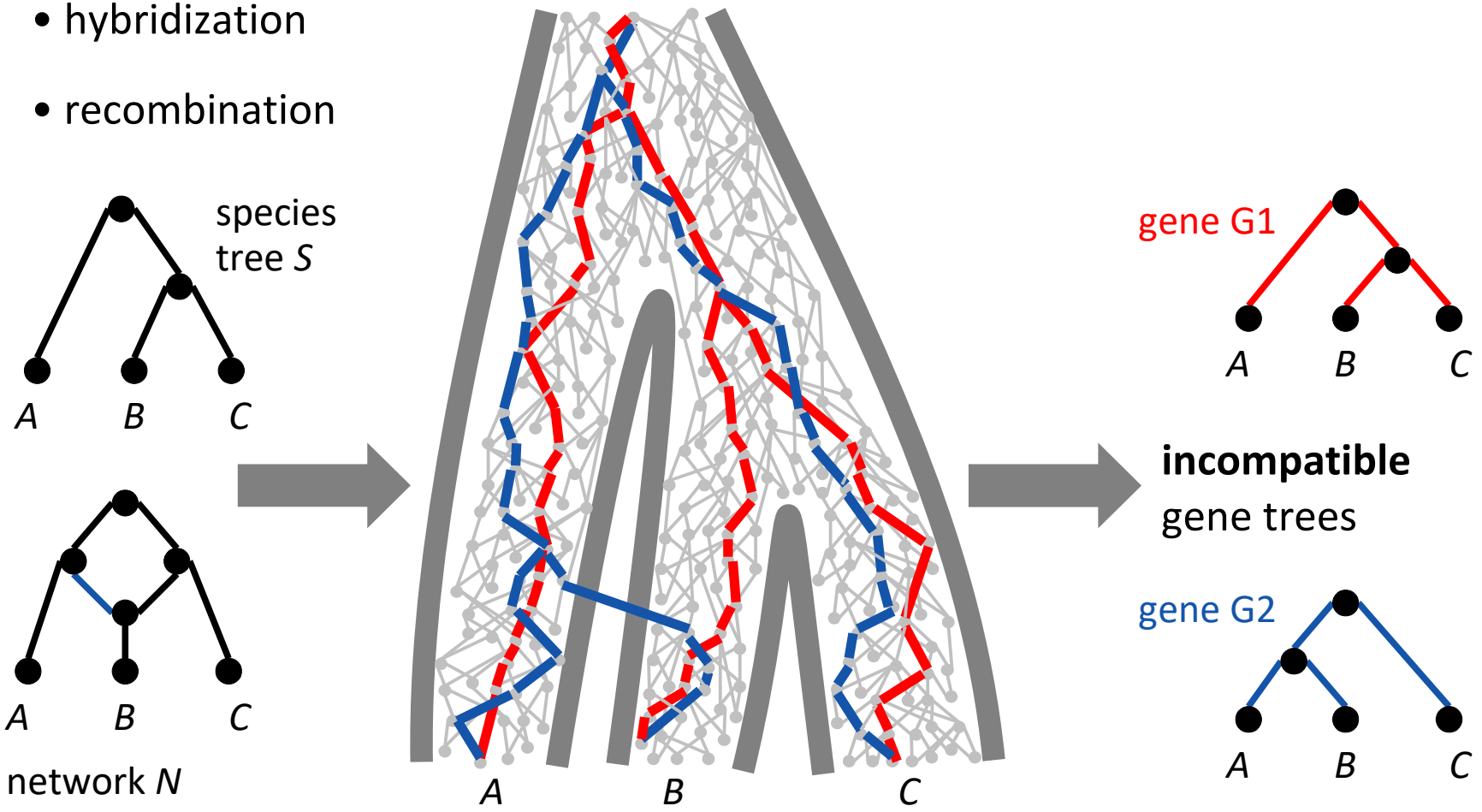
- lateral gene transfer
- hybridization
- recombination



Genetic material transfers

Transfers of genetic material between coexisting species:

- lateral gene transfer
- hybridization
- recombination

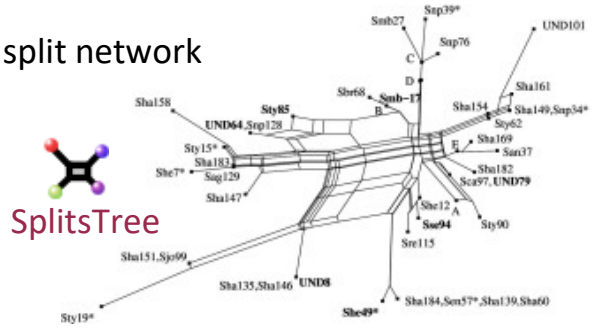


Phylogenetic networks

Phylogenetic network: network representing evolution data

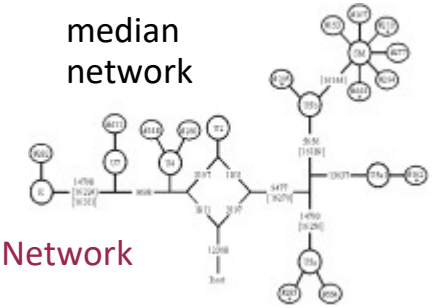
- **abstract / implicit / data-display** phylogenetic networks:
to **classify**, **visualize** **data**

split network



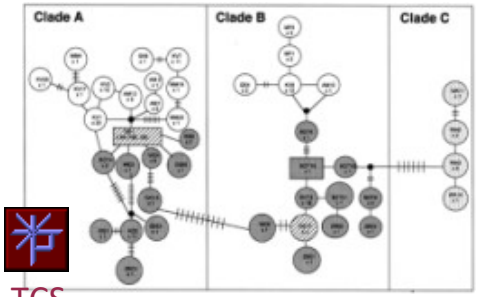
SplitsTree

median network



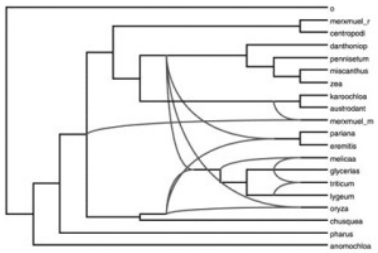
Network

minimum spanning network



TCS

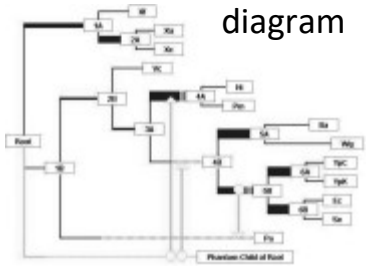
- **explicit** phylogenetic networks:
to **model** **evolution**



galled network

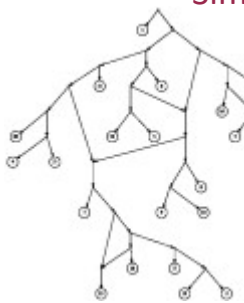
Dendroscope

synthesis diagram



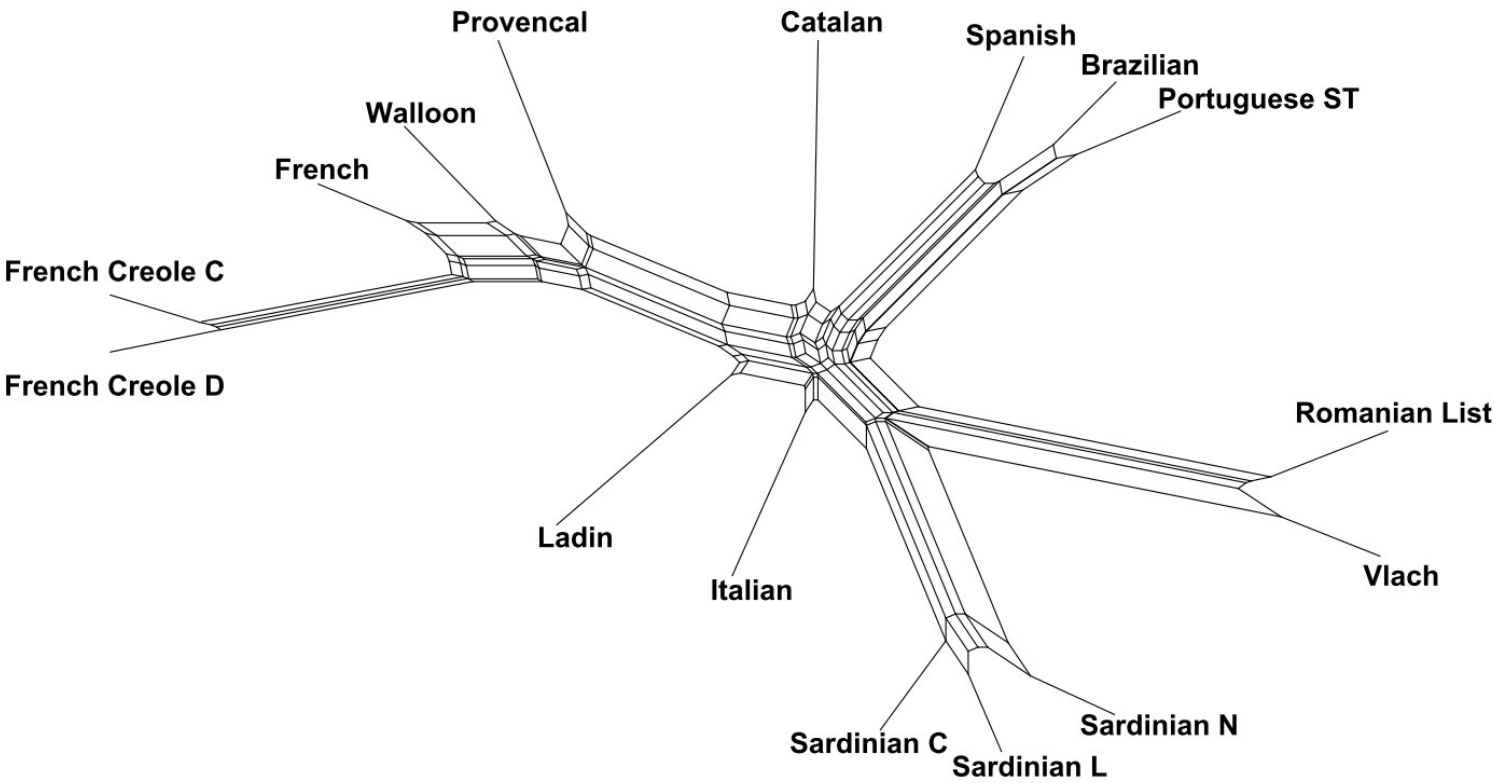
HorizStory

Simplistic



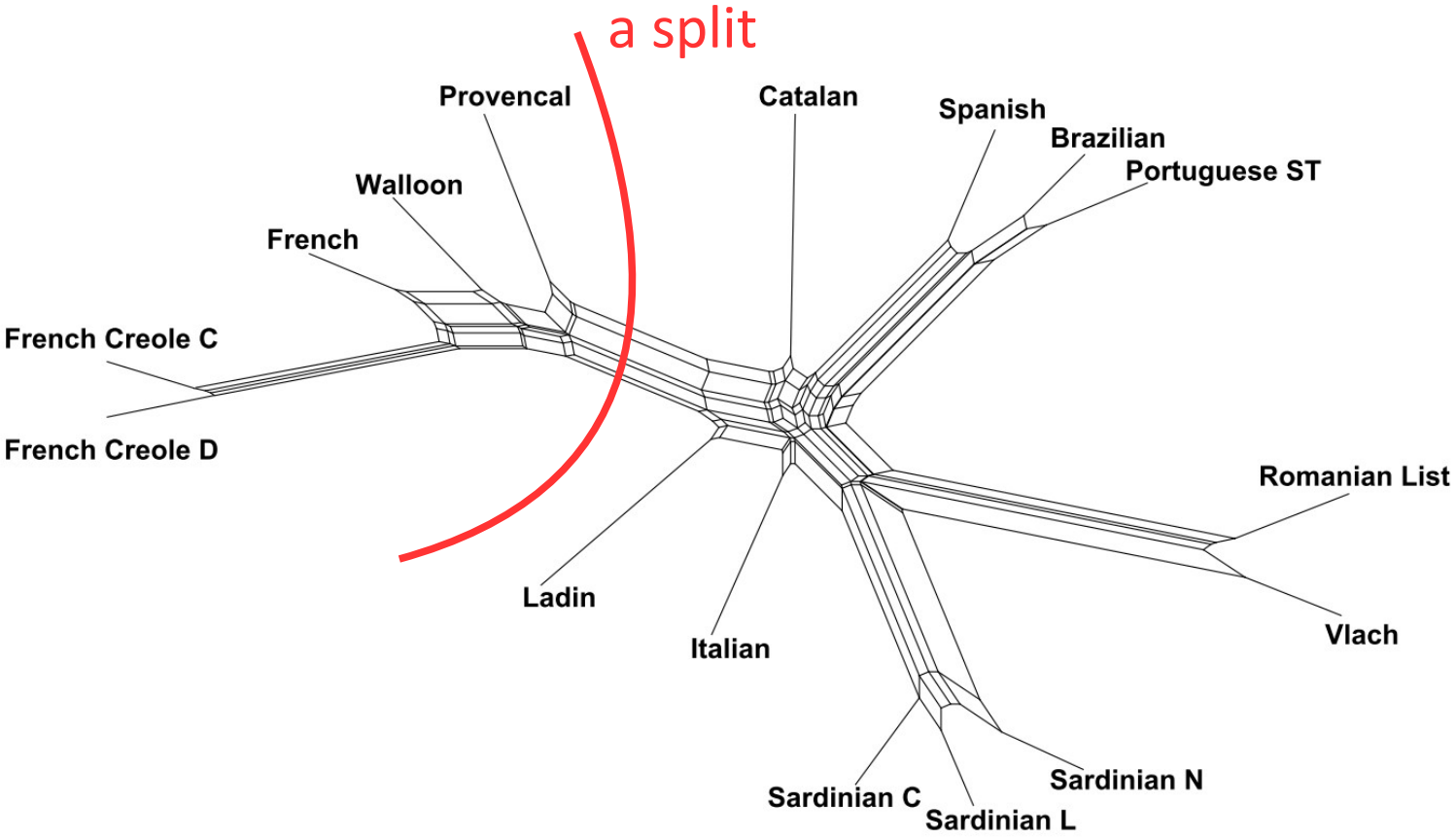
level-2 network

Abstract phylogenetic networks: split networks



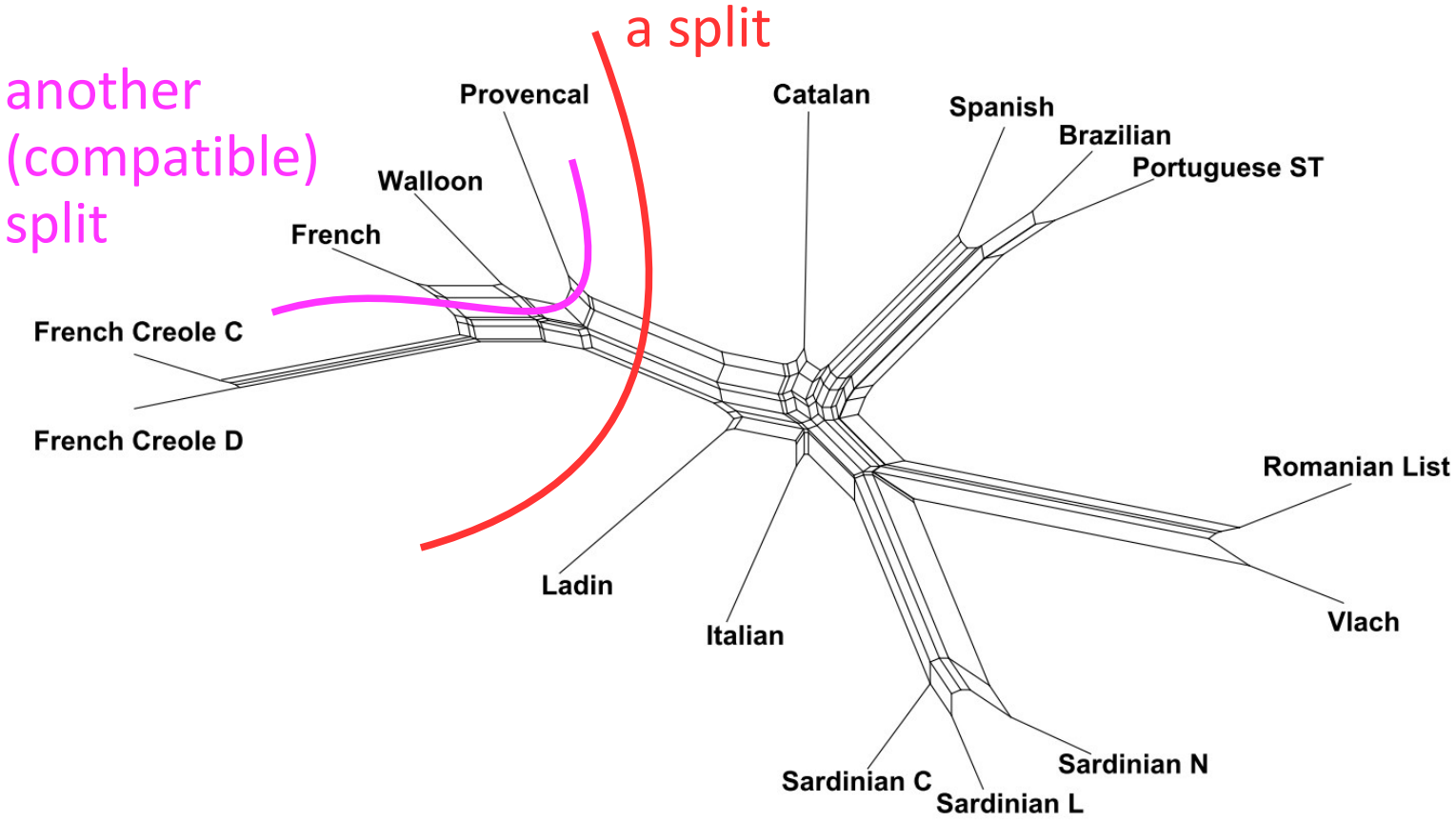
Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks



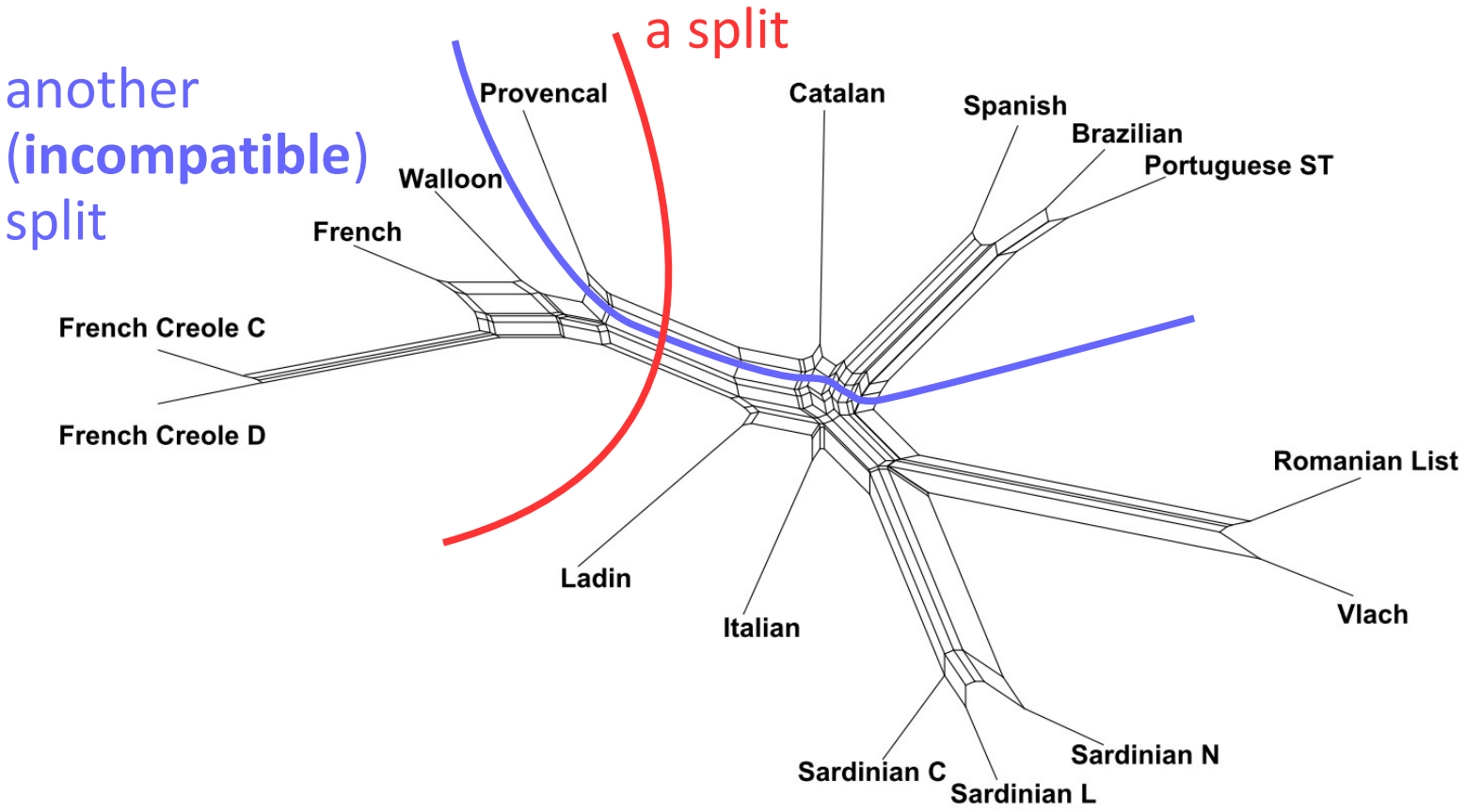
Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks



Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks

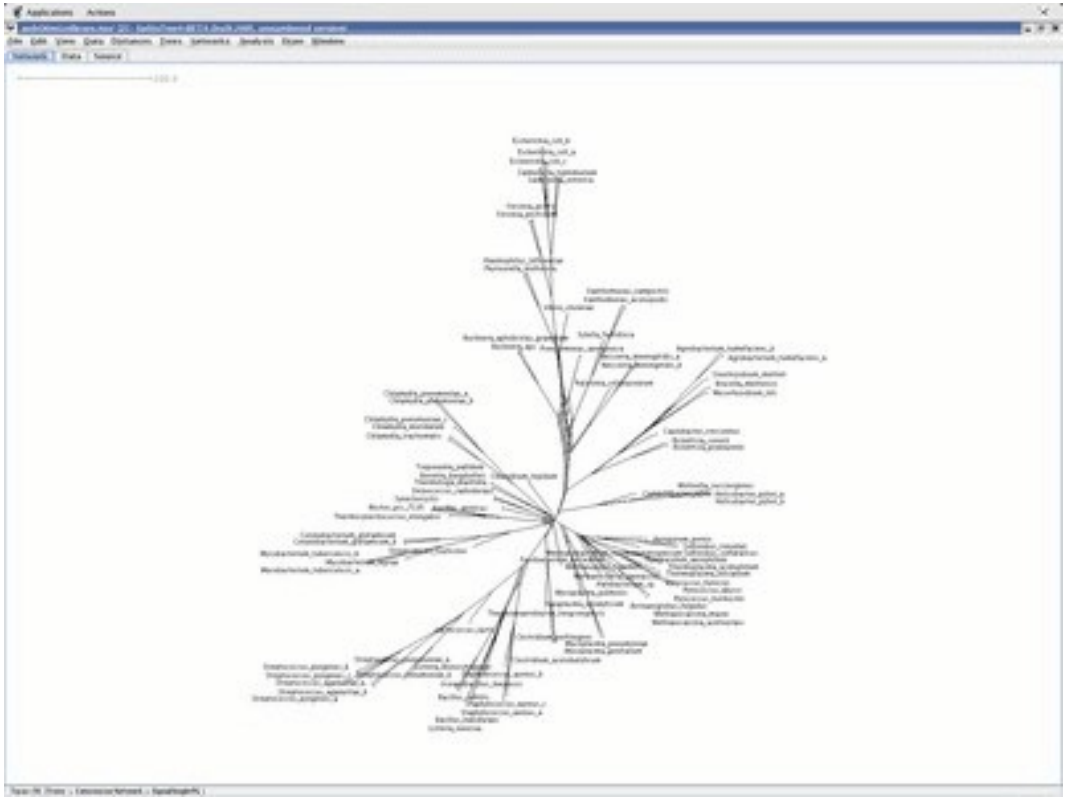


Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks

The “boxes” in split networks: incompatible splits

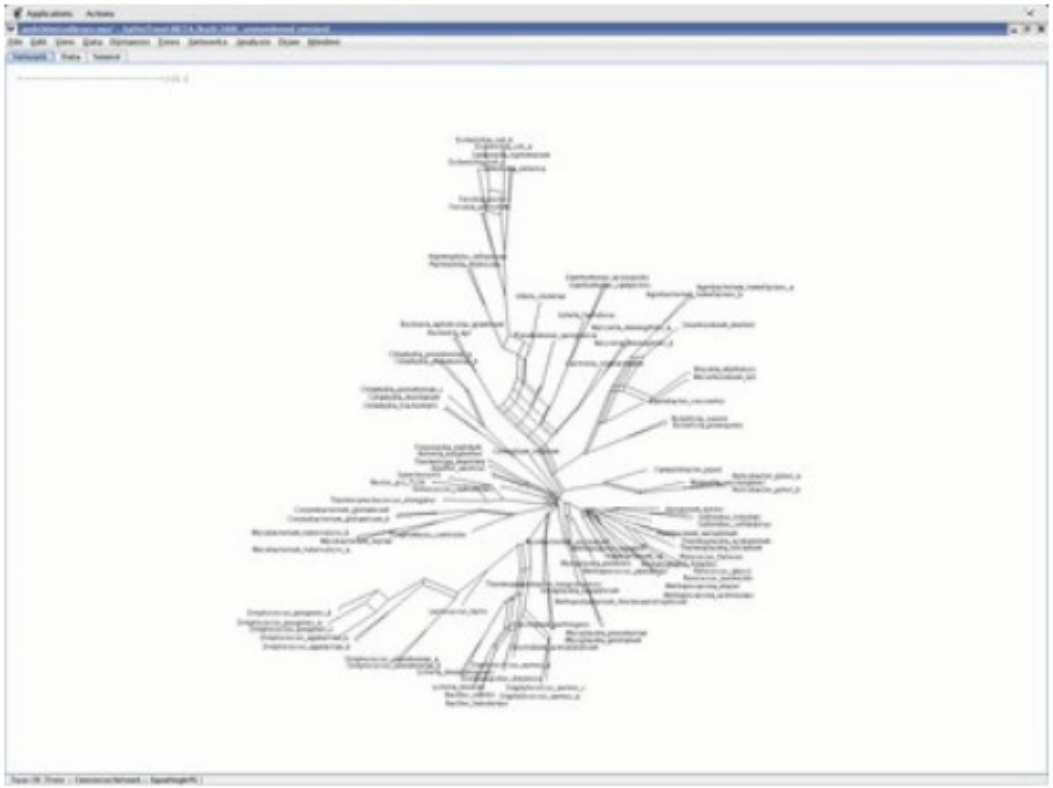
→ open them as much as possible in SplitsTree!



Abstract phylogenetic networks: split networks

The “boxes” in split networks: incompatible splits

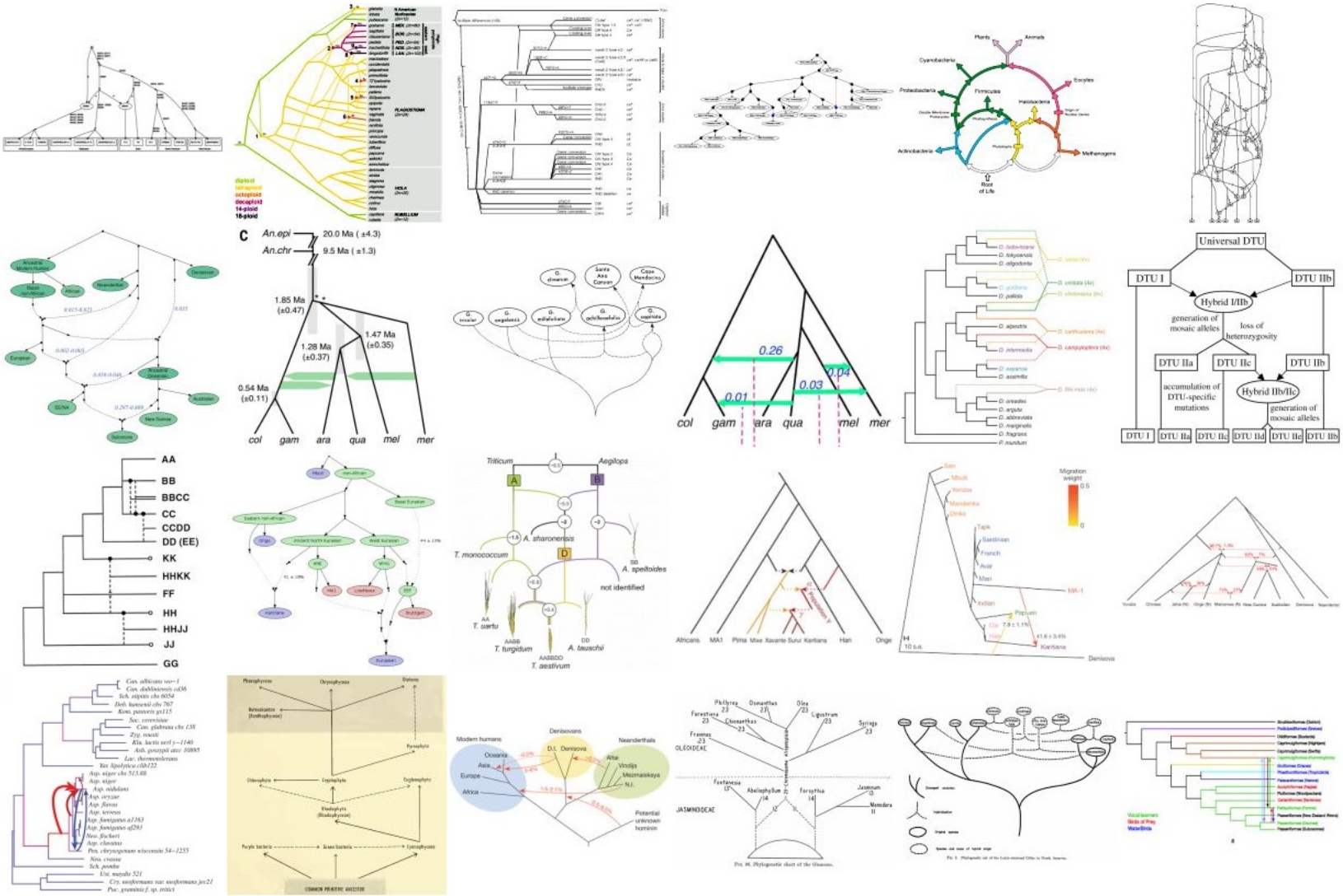
→ open them as much as possible in SplitsTree!



Philippe Gambette & Daniel H. Huson, Improved Layout of Phylogenetic Networks, IEEE/ACM TCBB 5(3), p. 472-479

Explicit phylogenetic networks

A gallery of explicit phylogenetic networks :

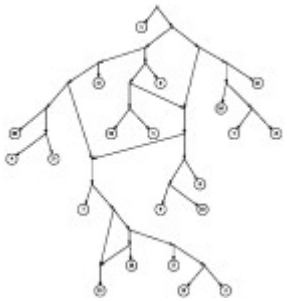


http://phylnet.univ-mlv.fr/recophync/networkDraw.php

Phylogenetic network reconstruction

```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2



{gene sequences}

distance methods

Bandelt & Dress 1992 - Legendre & Makarenkov 2000 - Bryant & Moulton 2002 - Chan, Jansson, Lam & Yiu 2006 - Willems, Tahiri & Makarenkov 2014

parsimony methods

Hein 1990 - Kececioglu & Gusfield 1994 - Jin, Nakhleh, Snir, Tuller 2009 - Park, Jin & Nakhleh 2010 - Kannan & Wheeler, 2012 - Kelk, Pardio, Scornavacca & van Iersel, 2017

likelihood methods

Snir & Tuller 2009 - Jin, Nakhleh, Snir, Tuller 2009 - Velasco & Sober 2009 - Meng & Kubatko 2009

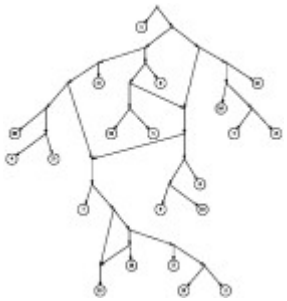
network *N*

Phylogenetic network reconstruction

Problem: methods are usually slow,
especially with rapidly increasing sequence length.

```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2



{gene sequences}

distance methods

Bandelt & Dress 1992 - Legendre & Makarenkov 2000 - Bryant & Moulton 2002 - Chan, Jansson, Lam & Yiu 2006 - Willems, Tahiri & Makarenkov 2014

parsimony methods

Hein 1990 - Kececioglu & Gusfield 1994 - Jin, Nakhleh, Snir, Tuller 2009 - Park, Jin & Nakhleh 2010 - Kannan & Wheeler, 2012 - Kelk, Pardio, Scornavacca & van Iersel, 2017

likelihood methods

Snir & Tuller 2009 - Jin, Nakhleh, Snir, Tuller 2009 - Velasco & Sober 2009 - Meng & Kubatko 2009

network *N*

Phylogenetic network reconstruction from trees

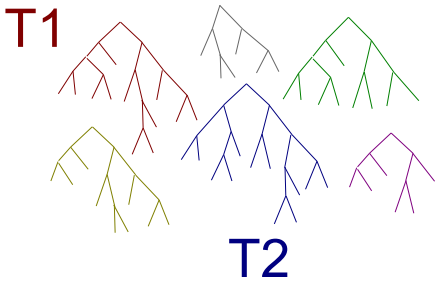
```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2

{gene sequences}

Reconstruction of a tree for each gene present in several species

Guindon & Gascuel, SB, 2003



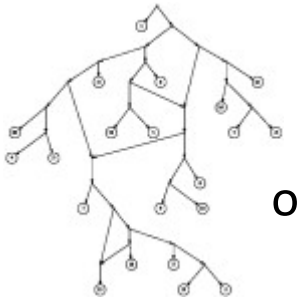
{trees}

HOGENOM Database

Dufayard, Duret, Penel, Gouy, Rechenmann & Perrière, BioInf, 2005

Tree reconciliation or consensus

explicit network



optimal super-network *N*:

- contains the input trees
- has the smallest number of reticulations

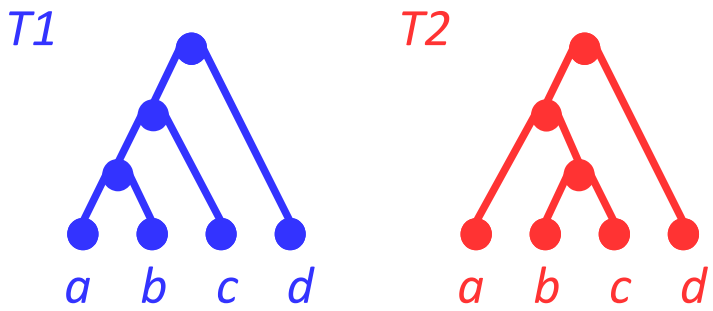
http://doua.prabi.fr/databases/hogenom/home.php?contents=hogenom4

Phylogenetic network reconstruction from trees

The **“hybridization network” problem**:

given 2 trees, find the smallest network containing both of them with the minimum number of hybrid vertices

Easy to find a network containing the two trees!

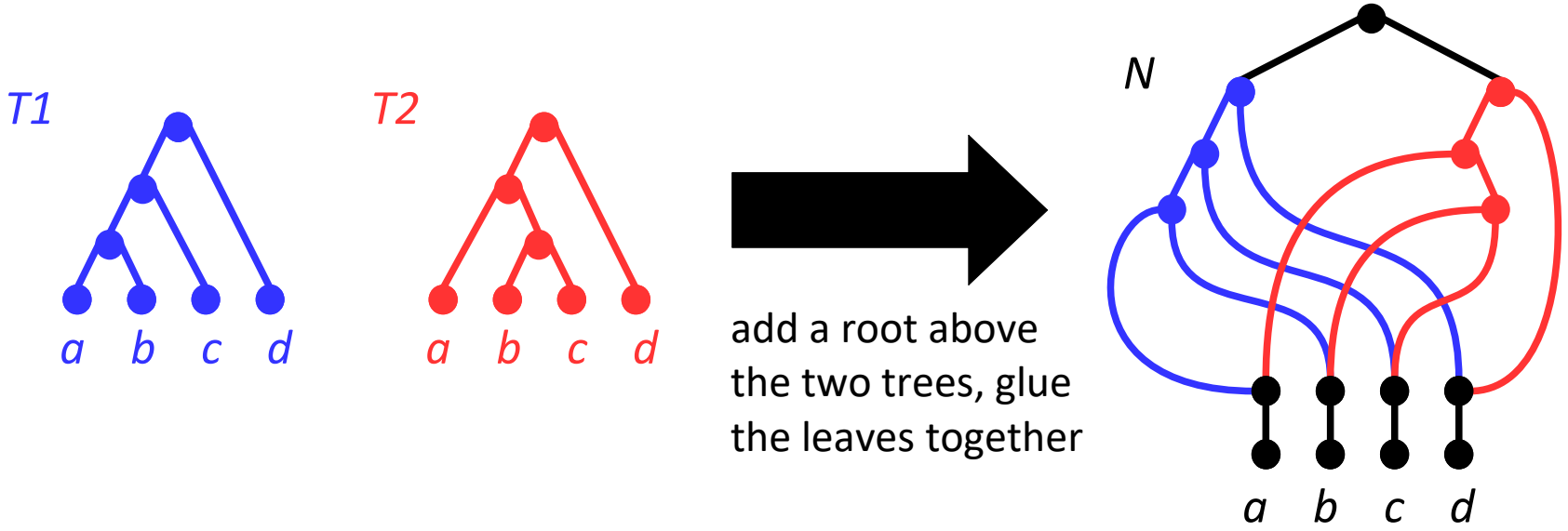


Phylogenetic network reconstruction from trees

The **"hybridization network" problem**:

given 2 trees, find the smallest network containing both of them with the minimum number of hybrid vertices

Easy to find a network containing the two trees!



But *n* hybrid vertices for trees with *n* leaves: not optimal!

Phylogenetic network reconstruction from trees

The **"hybridization network" problem**:

given 2 trees, find the smallest network containing both of them with the minimum number of hybrid vertices

NP-hard to minimize the number of hybrid vertices

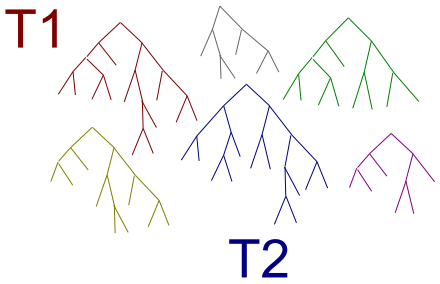
Bordewich & Semple (2007) Discrete Appl Math

Phylogenetic network reconstruction from trees

```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

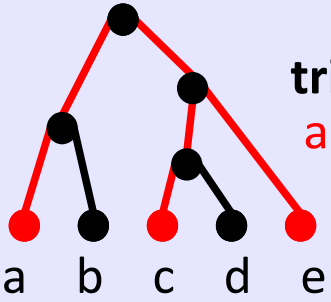
G1 G2

{gene sequences}



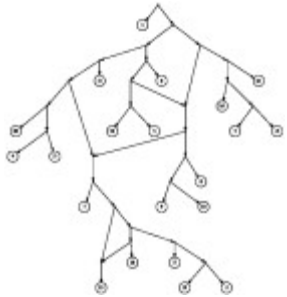
{trees}

{triplets}



triplet
a|ce

explicit network



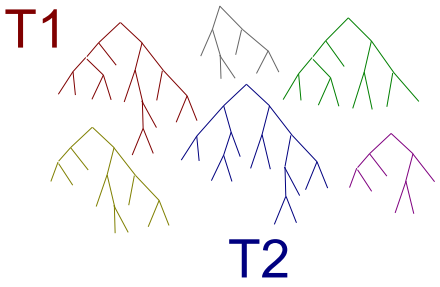
optimal super-network *N*:
- contains the input triplets
- has the smallest number of reticulations

Phylogenetic network reconstruction from trees

```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

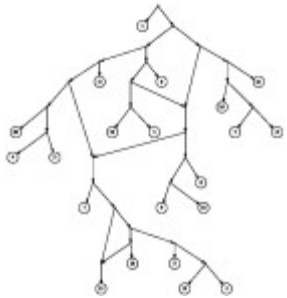
G1 G2

{gene sequences}



{trees}

{triplets}



explicit network

optimal super-network *N*:

- contains the input triplets
- has the smallest number of reticulations

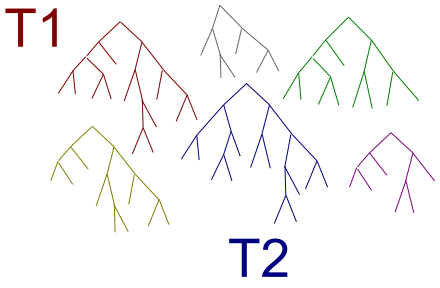
→ **still NP-hard**

Phylogenetic network reconstruction from trees

```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2

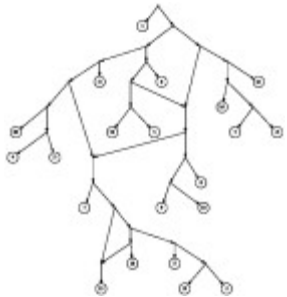
{gene sequences}



{trees}

{triplets} (“dense” set)

explicit network



super-network *N*:

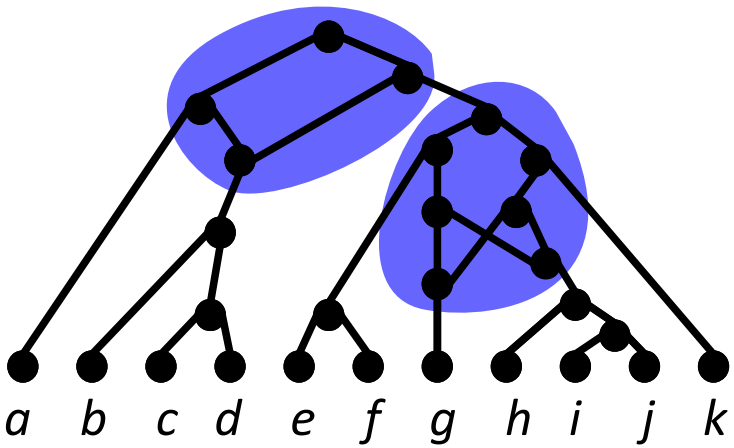
- contains the input triplets
- has level at most *k* (i.e. close to a tree)

→ polynomial time algorithm for fixed *k*:
Habib & To, CPM 2009, JBCB 2012

Outline

- Discovering graph classes with Michel
- An introduction to phylogenetic networks
- Classes of phylogenetic networks
- The Tree Containment Problem
- Counting phylogenetic networks

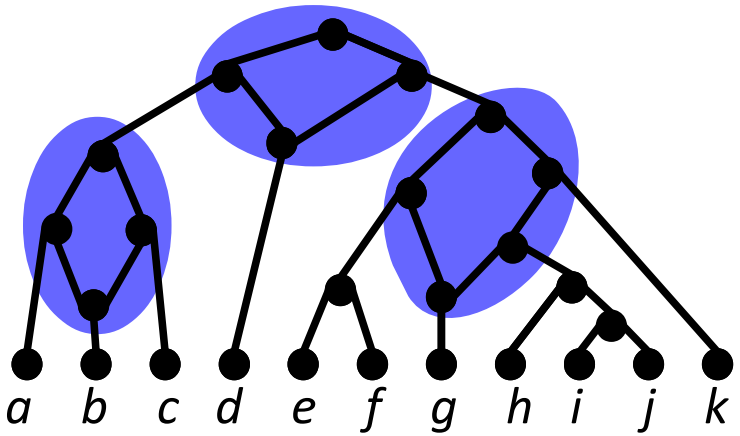
Classes of phylogenetic networks: level-k networks



level-2 network

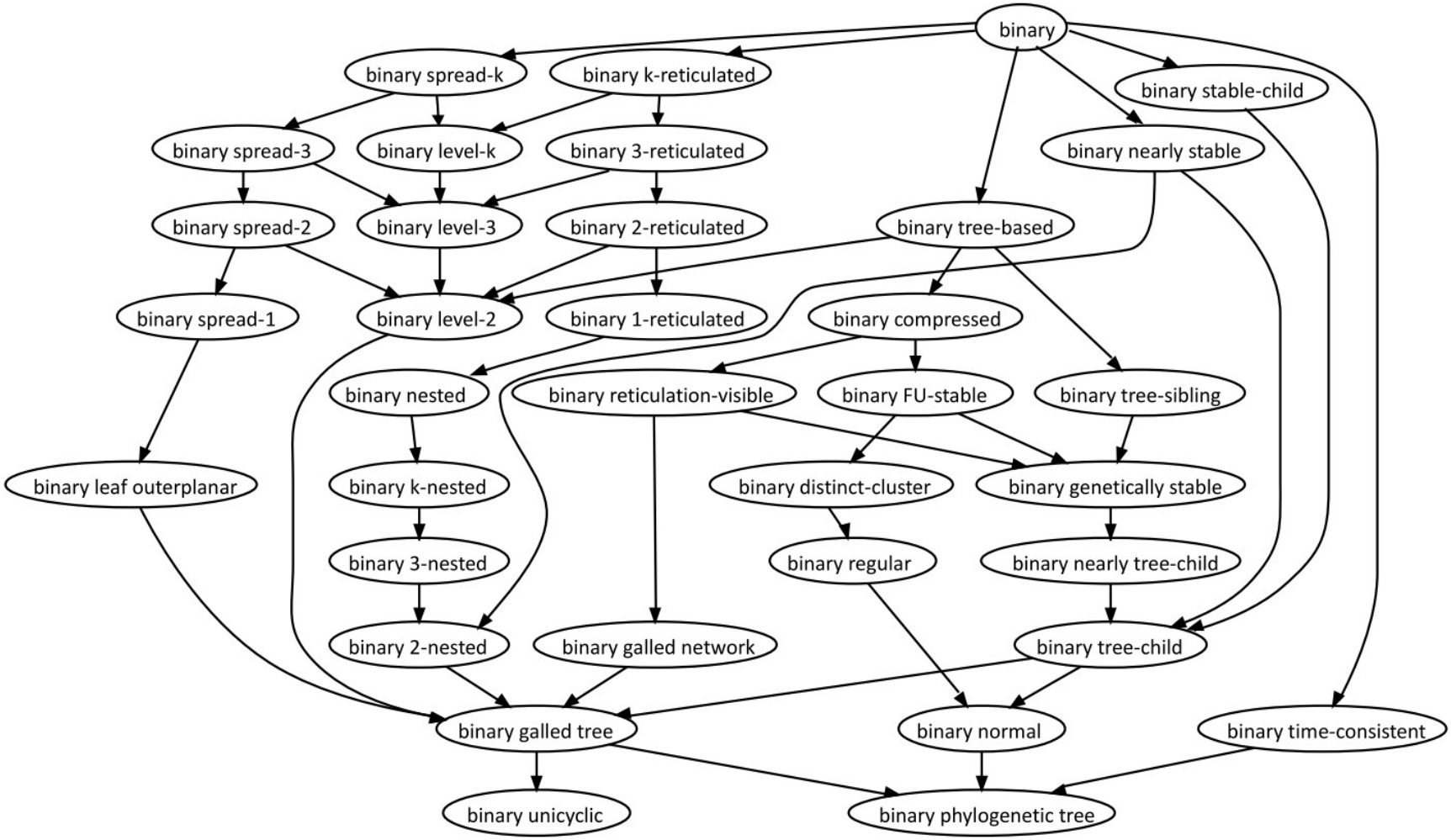
level =
maximum number of reticulations per **blob** (=bridgeless component of the underlying undirected graph).

level-1 network
(“*galled tree*”)



Classes of phylogenetic networks: ISIPhyNC

Information System on Inclusions of Phylogenetic Network Classes



joint work with Maxime Morgado and Narges Tavassoli

http://phylnet.univ-mlv.fr/isiphync/

Classes of phylogenetic networks: inclusions

binary	uni	gall	tree	nea	gall	gen	reti	tree	con	nor	regu	dist	FU-	nea	tree	nes	2-n	3-n	leve	leve	leaf	spr	spr	spr	tim
unicyclic	=	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5
galled tree	6	=	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5
tree-child	6	7	=	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	7	7	7	15	15	7	7			5
nearly tree-child	6	7	8	=	8	⊂	⊂	⊂	⊂	5	5	5	⊂	8	⊂	7	7	7	8	15	7	7			5
galled network	2	2	1	1	=	1	⊂	1	⊂	1	1	1	14	21	⊂	7	7	7	18	18	7	7			5
genetically stable	4	4	4	4	4	=	⊂	⊂	⊂	5	5	5	⊂	8	⊂	4	4	4	8	15	7	7			4
reticulation-visibl	2	2	1	1	4	1	=	1	⊂	1	1	1	14	8	⊂	4	4	4	8	15	7	7			4
tree-sibling	4	4	4	4	4	12	12	=	12	5	5	5	12	8	⊂	4	4	4	8	15	7	7			4
compressed	2	2	1	1	4	1	11	1	=	1	1	1	14	8	⊂	4	4	4	8	15	7	7			4
normal	7	7	⊂	⊂	13	⊂	⊂	⊂	⊂	=	⊂	⊂	⊂	⊂	⊂	7	7	7	15	15	7	7			22
regular	2	2	2	2	13	2	24	2	⊂	2	=	⊂	⊂	8	⊂	7	7	7	8	15	7	7			22
distinct-cluster	2	2	2	2	13	2	24	2	⊂	2	9	=	⊂	8	⊂	7	7	7	8	15	7	7			22
FU-stable	2	2	1	1	4	1	11	1	⊂	1	1	1	=	8	⊂	4	4	4	8	15	7	7			4
nearly stable	2	2	2	2	3	2	3	2	3	2	3	3	3	=	3	7	7	7	3	15	3	3			5
tree-based	2	2	1	1	4	1	12	1	12	1	1	1	12	8	=	4	4	4	8	15	7	7			4
nested	3	3	3	3	3	3	3	3	3	3	3	3	3	19	3	=	20	20	3	16	3	3			5
2-nested	3	3	3	3	3	3	3	3	3	3	3	3	3	⊂	3	⊂	=	⊂	3	16	3	3			5
3-nested	3	3	3	3	3	3	3	3	3	3	3	3	3	19	3	⊂	19	=	3	16	3	3			5
level-2	4	4	4	4	4	12	12	14	12	4	4	4	12	17	⊂	4	4	4	=	⊂	7	7	⊂	⊂	4
level-3	2	2	2	2	3	2	3	2	3	2	3	3	3	8	3	4	4	4	3	=	3	3		⊂	4
leaf outerplanar	4	4	4	4	4	10	10	10	10	4	4	4	10	8	10	4	4	4	8	16	=	⊂	⊂	⊂	4
spread 1	4	4	4	4	4	10	10	10	10	4	4	4	10	8	10	4	4	4	8	16	9	=	⊂	⊂	4
spread 2	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	3	16	3	3	=	⊂	4
spread 3	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	3	16	3	3		=	4
time-consistent	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	7	7	7	3	23	3	3			=

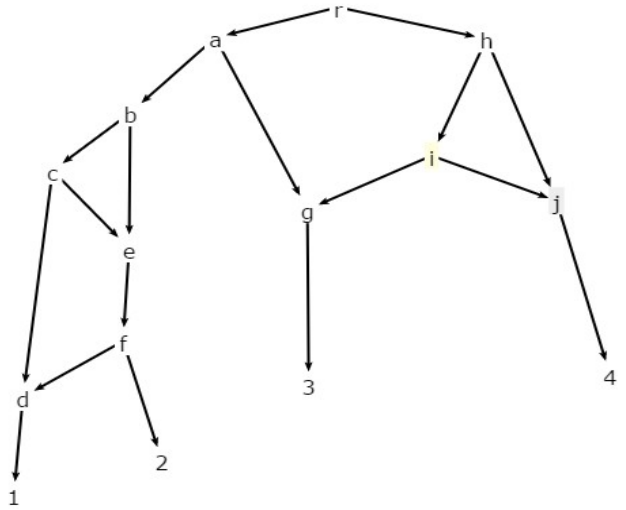
joint work with Maxime Morgado and Narges Tavassoli

Classes of phylogenetic networks: inclusions

binary	uni	gall	tree	nea	gall	gen	reti	tree	con	nor	regu	dist	FU-	nea	tree	nes	2-n	3-n	leve	leve	leaf	spr	spr	spr	tim
unicyclic	=	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5
galled tree	6	=	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5
tree-child	6	7	=	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	7	7	7	15	15	7	7			5
nearly tree-child	6	7	8	=	8	⊂	⊂	⊂	⊂	5	5	5	⊂	8	⊂	7	7	7	8	15	7	7			5
galled network	2	2	1	1	=	1	⊂	1	⊂	1	1	1	14	21	⊂	7	7	7	18	18	7	7			5
genetically stable	4	4	4	4	4	=	⊂	⊂	⊂	5	5	5	⊂	8	⊂	4	4	4	8	15	7	7			4
reticulation-visibl	2	2	1	1	4	1	=	1	⊂	1	1	1	14	8	⊂	4	4	4	8	15	7	7			4
tree-sibling	4	4	4	4	4	12	12	=	12	5	5	5	12	8	⊂	4	4	4	8	15	7	7			4
compressed	2	2	1	1	4	1	11	1	=	1	1	1	14	8	⊂	4	4	4	8	15	7	7			4
normal	7	7	⊂	⊂	13	⊂	⊂	⊂	⊂	=	⊂	⊂	⊂	⊂	⊂	7	7	7	15	15	7	7			22
regular	2	2	2	2	13	2	24	2	⊂	2	=	⊂	⊂	8	⊂	7	7	7	8	15	7	7			22
distinct-cluster	2	2	2	2	13	2	24	2	⊂	2	9	=	⊂	8	⊂	7	7	7	8	15	7	7			22
FU-stable	2	2	1	1	4	1	11	1	⊂	1	1	1	=	8	⊂	4	4	4	8	15	7	7			4
nearly stable	2	2	2	2	3	2	3	2	3	2	3	3	3	=	3	7	7	7	3	15	3	3			5
tree-based	2	2	1	1	4	1	12	1	12	1	1	1	12	8	=	4	4	4	8	15	7	7			4
nested	3	3	3	3	3	3	3	3	3	3	3	3	3	19	3	=	20	20	3	16	3	3			5
2-nested	3	3	3	3	3	3	3	3	3	3	3	3	3	⊂	3	⊂	=	⊂	3	16	3	3			5
3-nested	3	3	3	3	3	3	3	3	3	3	3	3	3	19	3	⊂	19	=	3	16	3	3			5
level-2	4	4	4	4	4	12	12	14	12	4	4	4	12	17	⊂	4	4	4	=	⊂	7	7	⊂	⊂	4
level-3	2	2	2	2	3	2	3	2	3	2	3	3	3	8	3	4	4	4	3	=	3	3		⊂	4
leaf outerplanar	4	4	4	4	4	10	10	10	10	4	4	4	10	8	10	4	4	4	8	16	=	⊂	⊂	⊂	4
spread 1	4	4	4	4	4	10	10	10	10	4	4	4	10	8	10	4	4	4	8	16	9	=	⊂	⊂	4
spread 2	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	3	16	3	3	=	⊂	4
spread 3	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	3	16	3	3		=	4
time-consistent	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	7	7	7	3	23	3	3			=

joint work with Maxime Morgado and Narges Tavassoli

Classes of phylogenetic networks: inclusions



Classes containing this network or not

Classes which contain this network (with direct proof)

- binary genetically stable: All reticulation vertices are stable (*e* for leaf 2, *d* for leaf 1, *g* for leaf 3, *j* for leaf 4) and all have at least one parent which is stable (*f*, parent of *d*, stable for leaf 2; *b*, parent of *e*, stable for leaves 1 and 2; *a*, parent of *g*, stable for leaves 1 and 2; *h*, parent of *j*, stable for leaf 4)
- binary nearly stable: All vertices except *c* and *i* (whose unique parent is in both cases stable) are stable.
- binary level-2: Easy to check.
- binary leaf outerplanar: Easy to check.
- binary FU-stable: Easy to check.

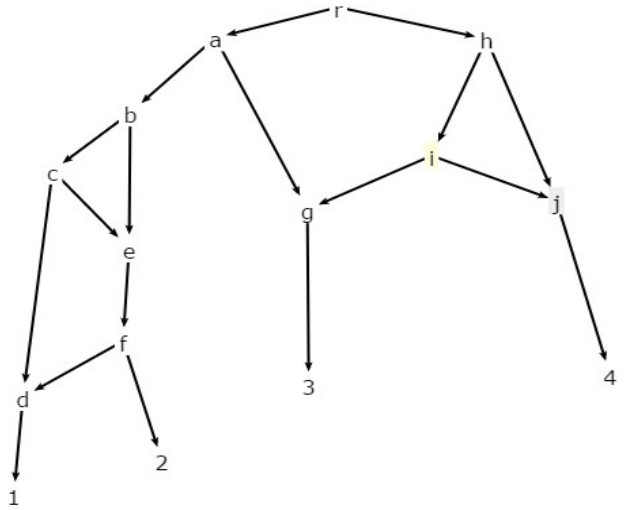
Classes which do not contain this network (with direct proof)

- binary nearly tree-child: Reticulation vertex *g* is stable but none of its parents *a* and *i* have the tree-path property.
- binary nested: Easy to check.
- binary time-consistent: The redundant arc from *b* to *e* makes it impossible to build a time-consistent labeling of the vertices.
- binary galled network: Reticulation arc (*c*,*e*) is a common arc of reticulation cycles from *b* to *e* and from *c* to *d*.
- binary distinct-cluster: Vertices *h* and *i* both have the same cluster: {3,4}
- binary 1-reticulated: Tree vertex *b* can reach 2 reticulation vertices by 2 directed internally vertex-disjoint paths: *d* and *e*.

joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/network.php?id=4

Classes of phylogenetic networks: inclusions

level = maximum number of reticulation vertices among all bridgeless components in the network
cluster = set of leaves below a vertex



Classes containing this network or not

Classes which contain this network (with direct proof)

- binary genetically stable: All reticulation vertices are stable (*e* for leaf 2, *d* for leaf 1, *g* for leaf 3, *j* for leaf 4) and all have at least one parent which is stable (*f*, parent of *d*, stable for leaf 2; *b*, parent of *e*, stable for leaves 1 and 2; *a*, parent of *g*, stable for leaves 1 and 2; *h*, parent of *j*, stable for leaf 4)
- binary nearly stable: All vertices except *c* and *i* (whose unique parent is in both cases stable) are stable.
- binary level-2: Easy to check.
- binary leaf outerplanar: Easy to check.
- binary FU-stable: Easy to check.

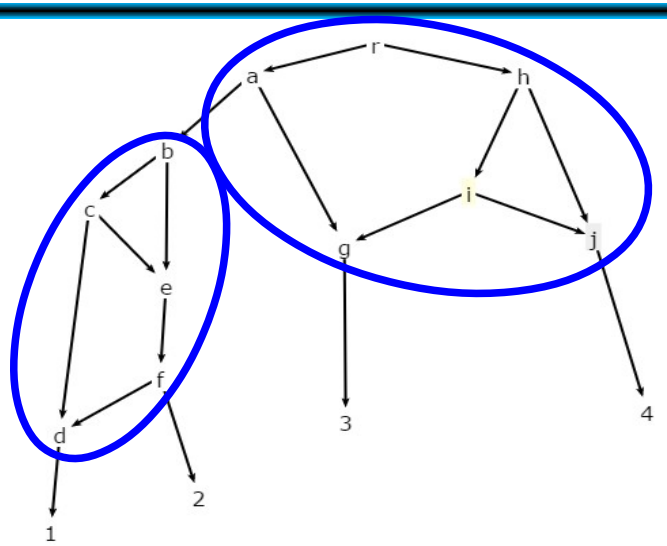
Classes which do not contain this network (with direct proof)

- binary nearly tree-child: Reticulation vertex *g* is stable but none of its parents *a* and *i* have the tree-path property.
- binary nested: Easy to check.
- binary time-consistent: The redundant arc from *b* to *e* makes it impossible to build a time-consistent labeling of the vertices.
- binary galled network: Reticulation arc (*c*,*e*) is a common arc of reticulation cycles from *b* to *e* and from *c* to *d*.
- binary distinct-cluster: Vertices *h* and *i* both have the same cluster: {3,4}
- binary 1-reticulated: Tree vertex *b* can reach 2 reticulation vertices by 2 directed internally vertex-disjoint paths: *d* and *e*.

joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/network.php?id=4

Classes of phylogenetic networks: inclusions

level = maximum number of reticulation vertices among all **bridgeless components** in the network
cluster = set of leaves below a vertex



Classes containing this network or not

Classes which contain this network (with direct proof)

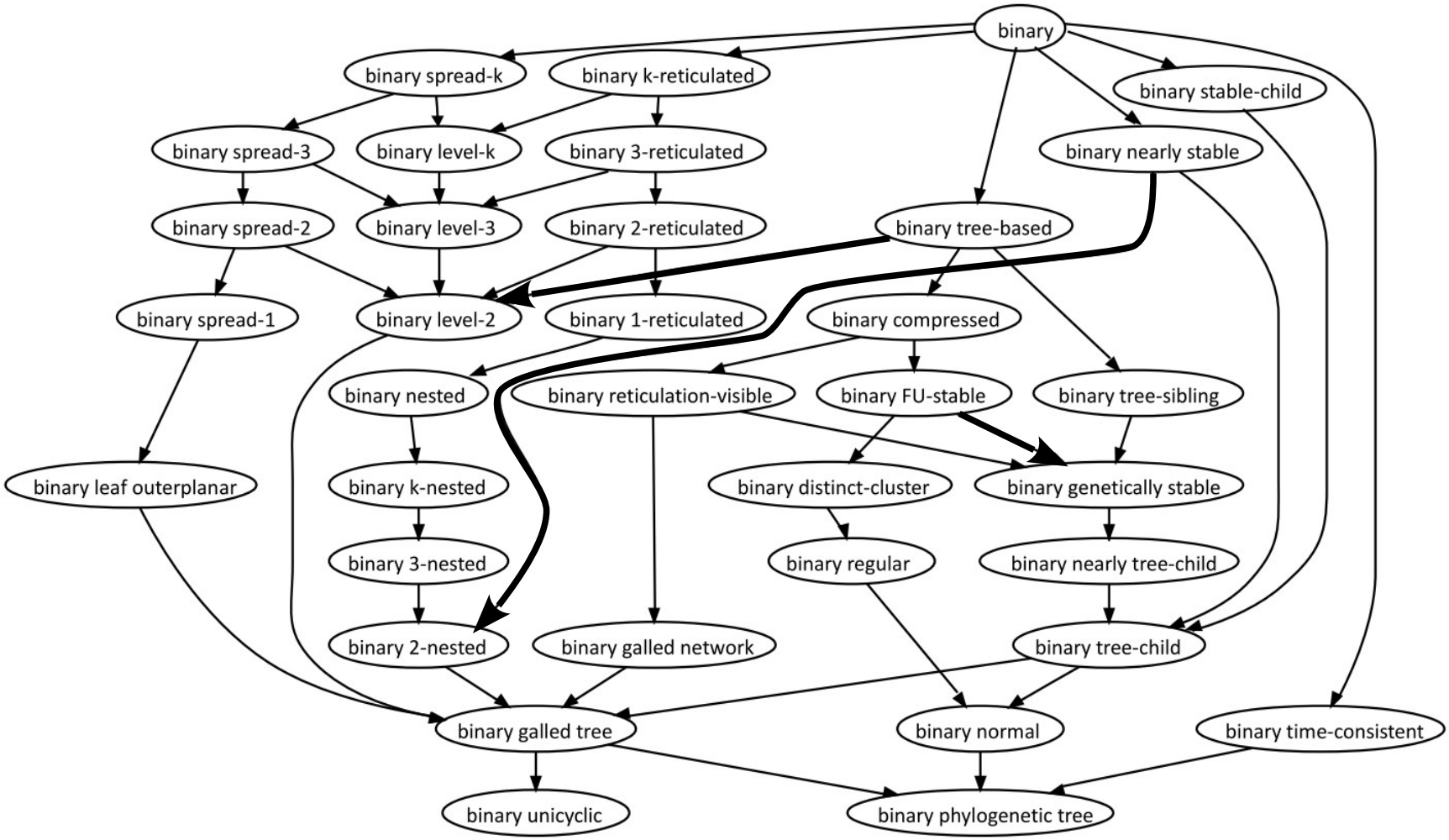
- binary genetically stable: All reticulation vertices are stable (*e* for leaf 2, *d* for leaf 1, *g* for leaf 3, *j* for leaf 4) and all have at least one parent which is stable (*f*, parent of *d*, stable for leaf 2; *b*, parent of *e*, stable for leaves 1 and 2; *a*, parent of *g*, stable for leaves 1 and 2; *h*, parent of *j*, stable for leaf 4)
- binary nearly stable: All vertices except *c* and *i* (whose unique parent is in both cases stable) are stable.
- binary level-2: Easy to check.
- binary leaf outerplanar: Easy to check.
- binary FU-stable: Easy to check.

Classes which do not contain this network (with direct proof)

- binary nearly tree-child: Reticulation vertex *g* is stable but none of its parents *a* and *i* have the tree-path property.
- binary nested: Easy to check.
- binary time-consistent: The redundant arc from *b* to *e* makes it impossible to build a time-consistent labeling of the vertices.
- binary galled network: Reticulation arc (*c*,*e*) is a common arc of reticulation cycles from *b* to *e* and from *c* to *d*.
- binary distinct-cluster: Vertices *h* and *i* both have the same cluster: {3,4}
- binary 1-reticulated: Tree vertex *b* can reach 2 reticulation vertices by 2 directed internally vertex-disjoint paths: *d* and *e*.

joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/network.php?id=4

Classes of phylogenetic networks: new inclusions



joint work with Maxime Morgado and Narges Tavassoli

Classes of phylogenetic networks: problems

Problems and properties studied on these classes

Problems

- **Tree Containment**: Does the input network contain the input tree on the same set of leaves?
- **Cluster Containment**: Does the input network contain the input subset of leaves as a softwired cluster?
- **Phylogenetic Network Isomorphism**: Are the two input phylogenetic networks isomorphic?

Properties

- **Upper bound on the number of vertices**: The number of vertices is bounded by the number of leaves.
- **Unbounded number of vertices**: The number of vertices is not bounded by the number of leaves.
- **Completeness for reconstruction from trees**: There exists a network in this class which displays all binary rooted phylogenetic trees on *n* leaves.

Database content

73 classes of phylogenetic networks including 35 classes of binary phylogenetic networks (defined in a total of 20 bibliographic references), 51 inclusion relationships proved directly between classes (including some found in a total of 9 bibliographic references), 21 networks (65 memberships to a class, 53 non-memberships to a class), 3 problems considered, 3 properties considered, 33 theorems proved directly (including some found in a total of 15 bibliographic references) including 23 positive results (which can be extended to subclasses) and 10 negative results (which can be extended to superclasses).

joint work with Maxime Morgado and Narges Tavassoli

Classes of phylogenetic networks: problems

Problems and properties studied on these classes

Problems

- **Tree Containment**: Does the input network contain the input tree on the same set of leaves?
- **Cluster Containment**: Does the input network contain the input subset of leaves as a softwired cluster?
- **Phylogenetic Network Isomorphism**: Are the two input phylogenetic networks isomorphic?

Properties

- **Upper bound on the number of vertices**: The number of vertices is bounded by the number of leaves.
- **Unbounded number of vertices**: The number of vertices is not bounded by the number of leaves.
- **Completeness for reconstruction from trees**: There exists a network in this class which displays all binary rooted phylogenetic trees on *n* leaves.

Database content

73 classes of phylogenetic networks including 35 classes of binary phylogenetic networks (defined in a total of 20 bibliographic references), 51 inclusion relationships proved directly between classes (including some found in a total of 9 bibliographic references), 21 networks (65 memberships to a class, 53 non-memberships to a class), 3 problems considered, 3 properties considered, 33 theorems proved directly (including some found in a total of 15 bibliographic references) including 23 positive results (which can be extended to subclasses) and 10 negative results (which can be extended to superclasses).

Problem **easy to solve** on class A → **easy to solve** on **subclass** B
Hard to solve on class B → **hard to solve** on **superclass** A
(similar to ISGCI)

joint work with Maxime Morgado and Narges Tavassoli

Outline

- Discovering graph classes with Michel
- Classes of phylogenetic networks
- An introduction to phylogenetic networks
- The Tree Containment Problem
- Counting phylogenetic networks

Phylogenetic network reconstruction from trees

Recall the **“hybridization network” problem**:

given 2 trees, find the smallest network containing both of them with the minimum number of hybrid vertices

Even **checking a solution** is **hard**!

The Tree Containment Problem (T.C.P.)

Input: A binary phylogenetic network *N* and

a tree *T* over the same set of taxa.

Question: Does *N* display *T*?

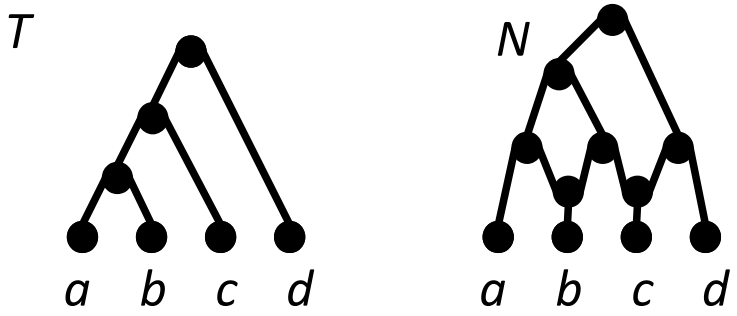
The Tree Containment Problem (T.C.P.)

Input: A binary phylogenetic network *N* and

a tree *T* over the same set of taxa.

Question: Does *N* display *T*?

→ Can we remove one incoming arc, for each vertex with >1 parent in *N*, such that the obtained tree is equivalent to *T*?



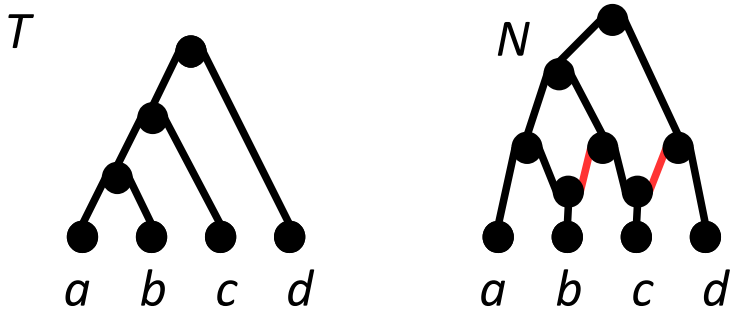
The Tree Containment Problem (T.C.P.)

Input: A binary phylogenetic network *N* and

a tree *T* over the same set of taxa.

Question: Does *N* display *T*?

→ Can we **remove one incoming arc**, for each vertex with >1 parent in *N*, such that the obtained tree is equivalent to *T*?



The Tree Containment Problem (T.C.P.)

Input: A binary phylogenetic network *N* and

a tree *T* over the same set of taxa.

Question: Does *N* display *T*?

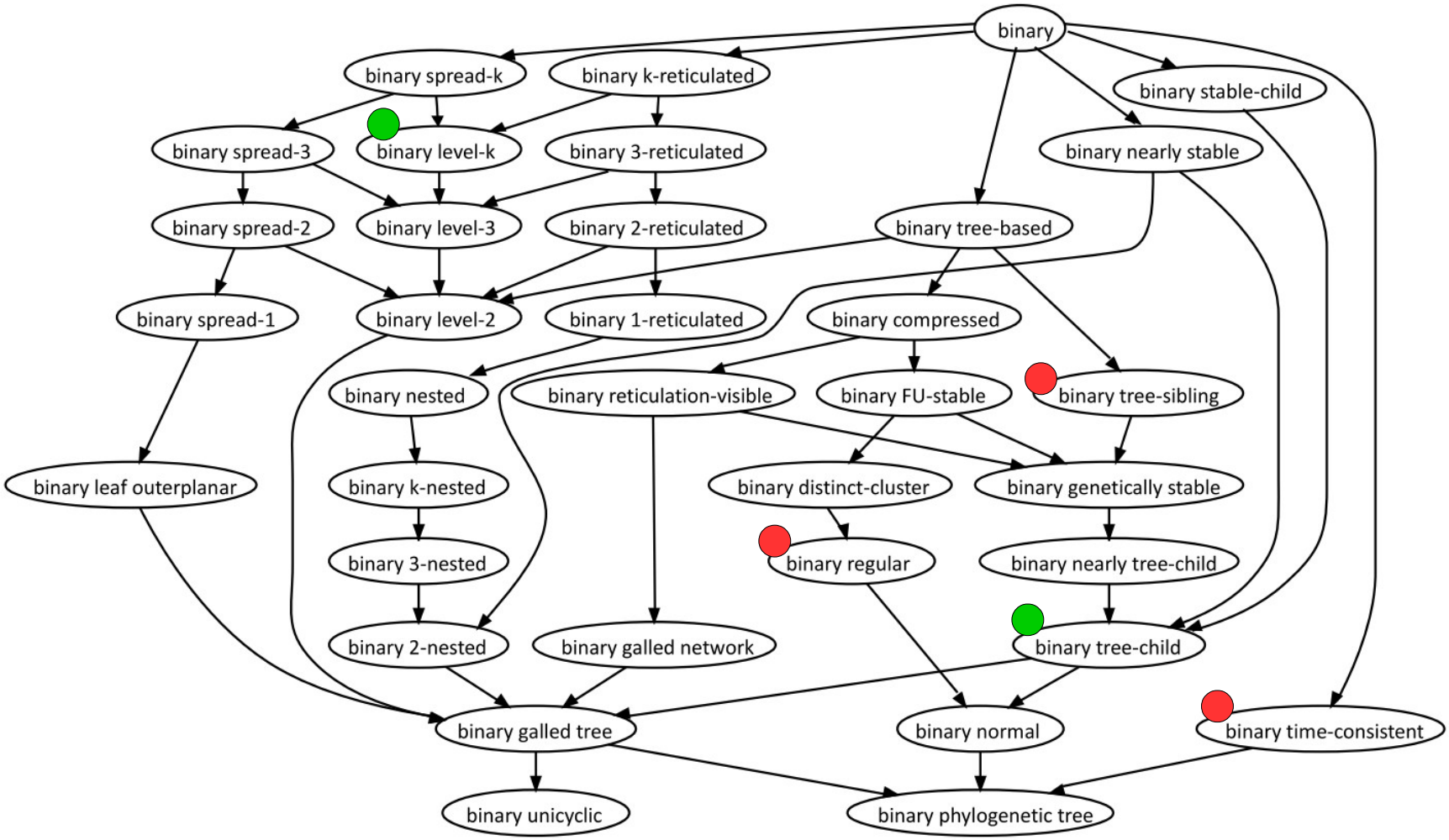
- **NP-complete** in general (Kanj, Nakhleh, Than & Xia, 2008)

- **NP-complete** for tree-sibling, time-consistent, regular networks (Iersel, Semple & Steel, 2010)

- **Polynomial-time solvable** for normal networks, for binary tree-child networks, and for level-k networks (Iersel, Semple & Steel, 2010)

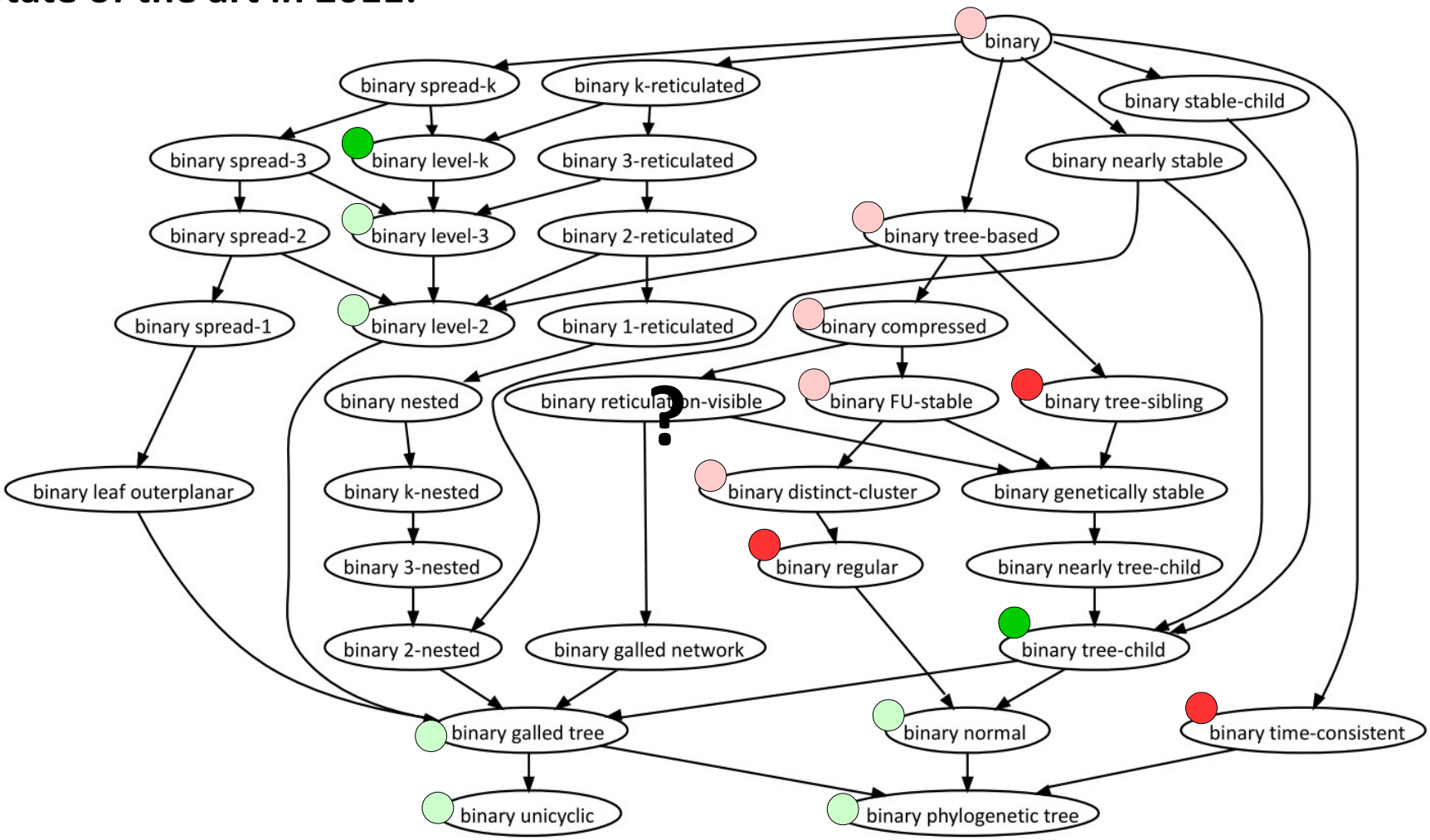
Classes of phylogenetic networks and the T.C.P.

State of the art in 2011:

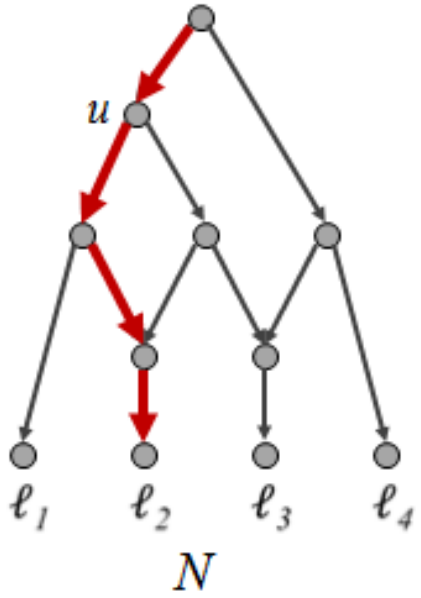


Classes of phylogenetic networks and the T.C.P.

State of the art in 2011:

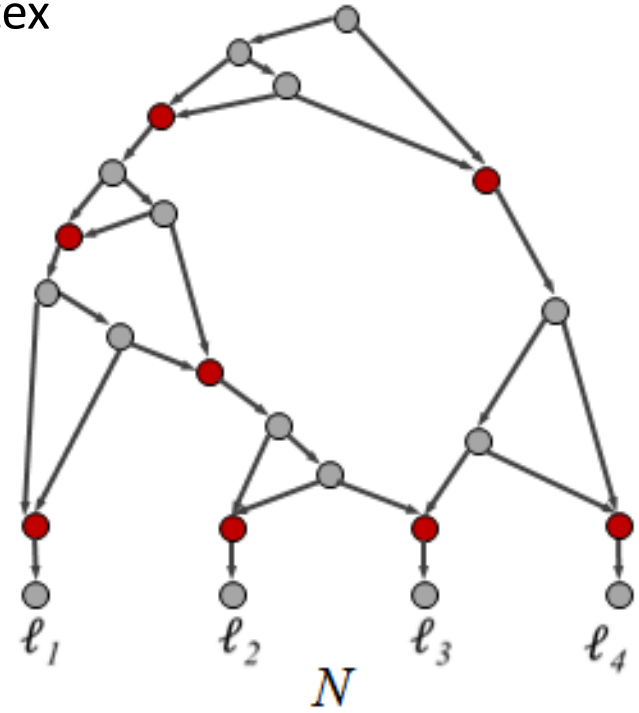


Reticulation-visible and nearly-stable networks



A vertex *u* is **stable** if there exists a leaf *l* such that all paths from the root to *l* go through *u*.

A phylogenetic network is **reticulation-visible** if every reticulation vertex is stable.

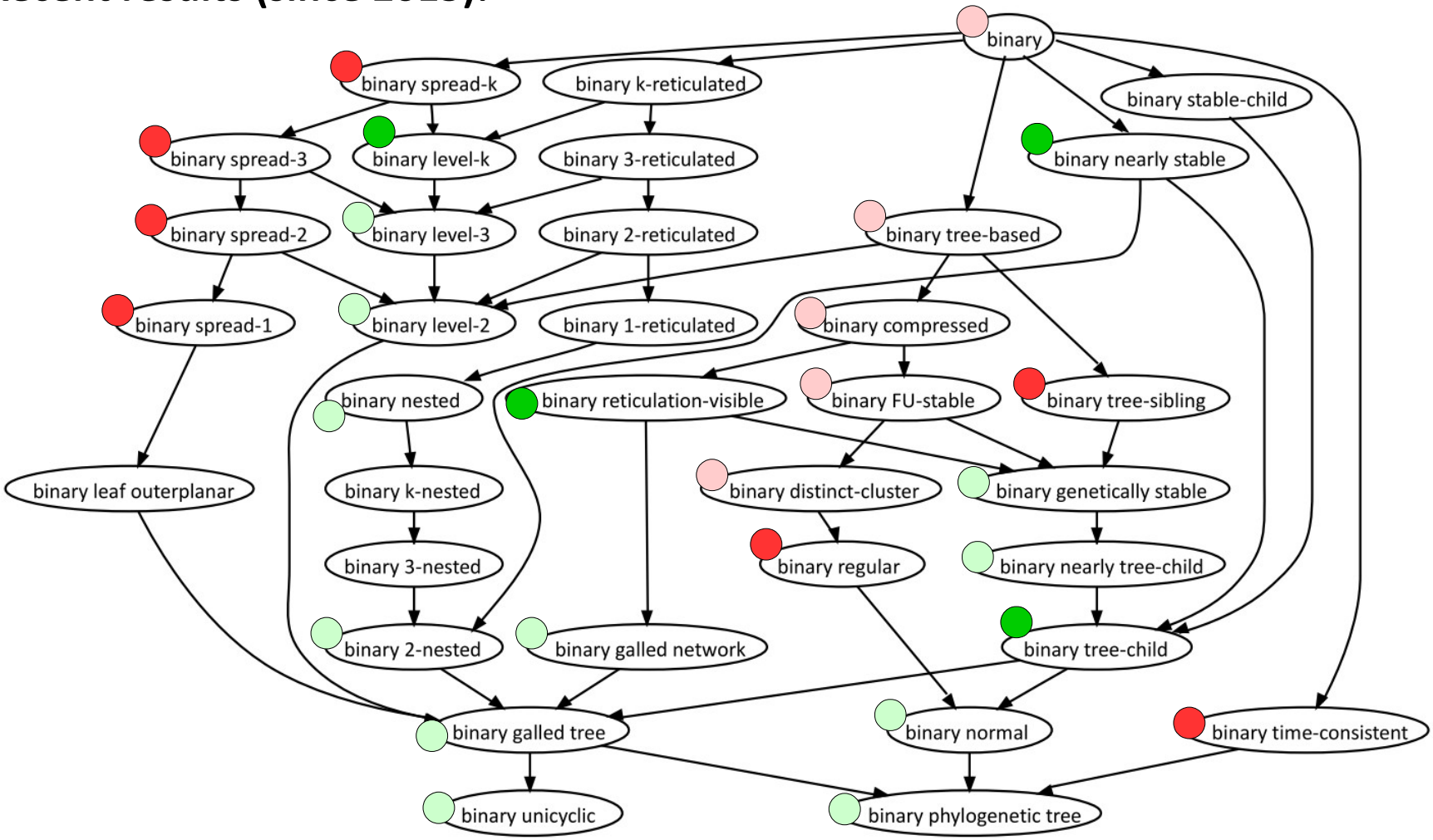


A phylogenetic network is **nearly-stable** if for each vertex, either it is stable or its parents are.

Gambette, Gunawan, Labarre, Vialette & Zhang, RECOMB 2015

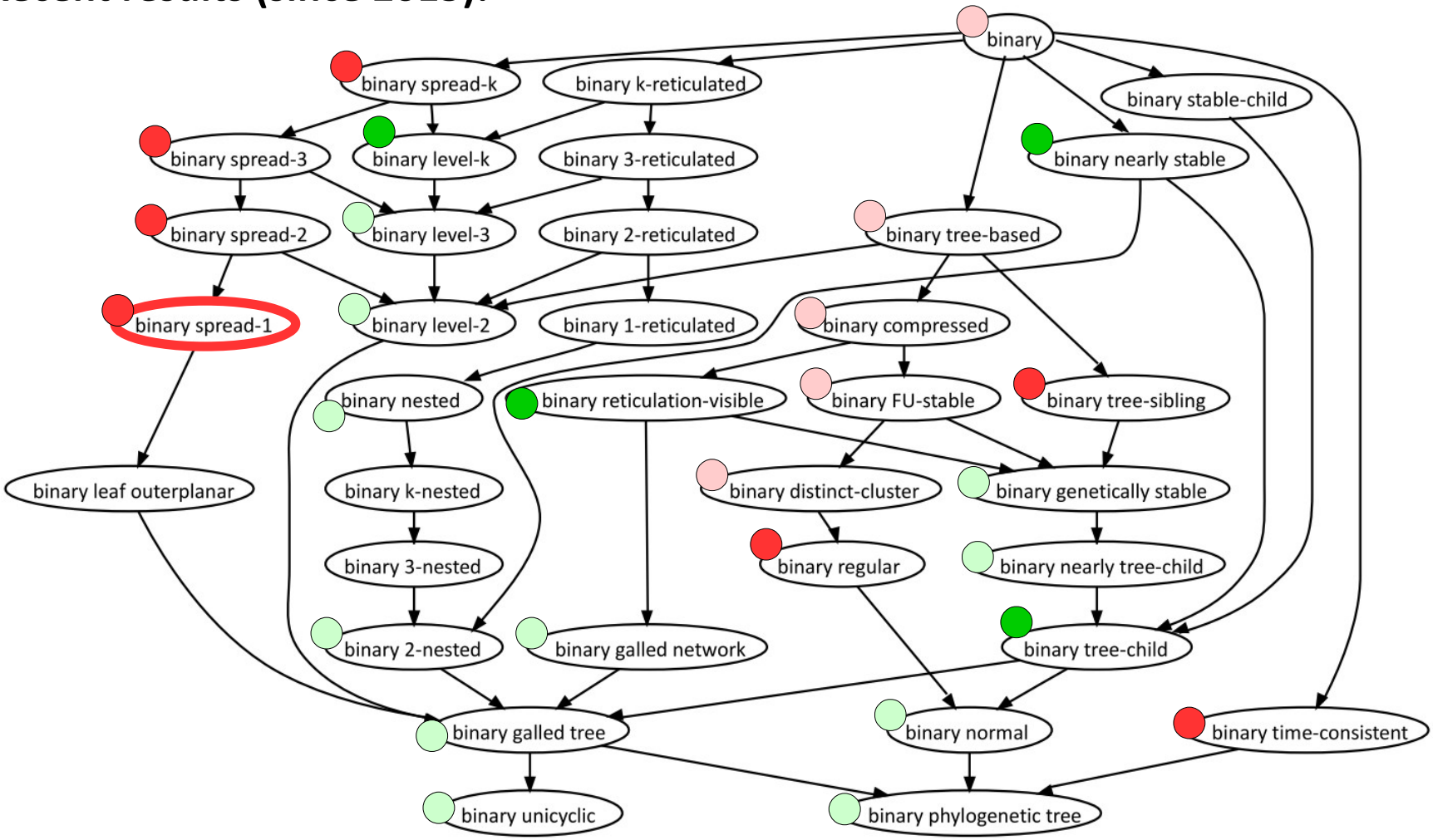
Classes of phylogenetic networks and the T.C.P.

Recent results (since 2015):



Classes of phylogenetic networks and the T.C.P.

Recent results (since 2015):



Spread-k networks: leaves can be ordered such that the set of leaves below each vertex is a union of **at most *k* intervals**.

Fête de la Science at UPEM (science festival)

Photos : Campus numérique de l'UPEM

Outline

- Discovering graph classes with Michel
- Classes of phylogenetic networks
- An introduction to phylogenetic networks
- The Tree Containment Problem
- Counting phylogenetic networks

Counting labeled unrooted level-1 networks

Unrooted level-1 networks:
explicit formula for *n* leaves, *c* cycles, *m* edges involved in the cycles.

Semple & Steel, *TCBB*, 2006

Counting labeled unrooted level-1 networks

Unrooted level-1 networks:
explicit formula for *n* leaves, *c* cycles, *m* edges involved in the cycles.

Pointing + bijection:
Bijection between labeled unrooted level-1 networks with *n*+1 leaves and labeled pointed level-1 networks with *n* leaves.

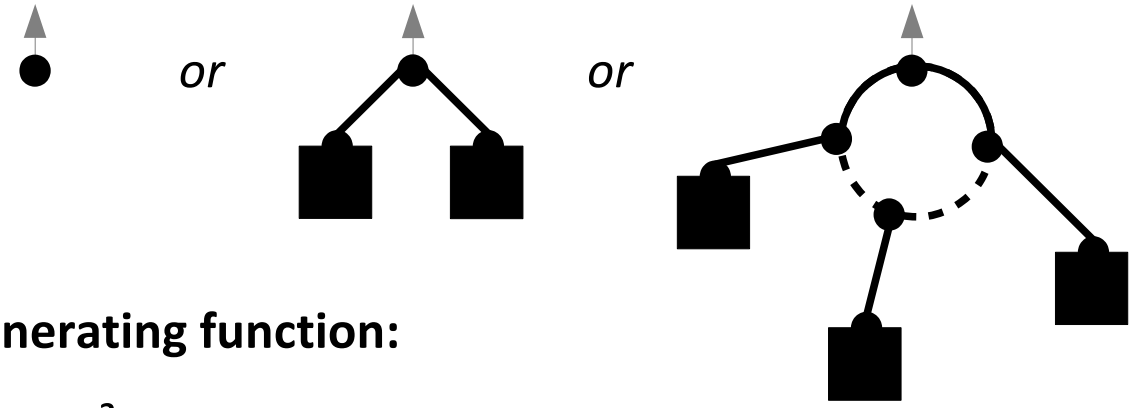
Semple & Steel, *TCBB*, 2006

Counting labeled unrooted level-1 networks

Unrooted level-1 networks:
explicit formula for *n* leaves, *c* cycles, *m* edges involved in the cycles.

Pointing + bijection:
Bijection between labeled unrooted level-1 networks with *n*+1 leaves and labeled pointed level-1 networks with *n* leaves.

Recursive decomposition of pointed level-1 networks with *n* leaves:



Exponential generating function:

$$G = z + \frac{1}{2}G^2 + \frac{1}{2}\frac{G^2}{(1-G)}$$

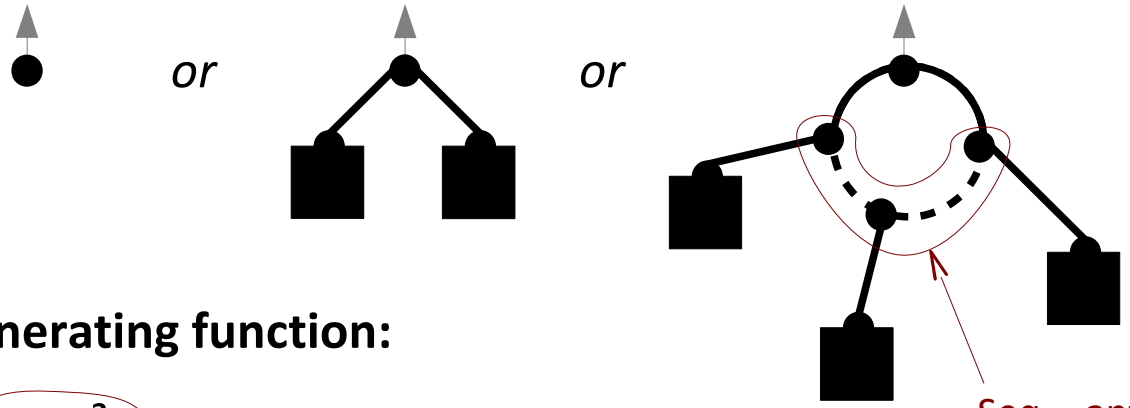
Semple & Steel, *TCBB*, 2006

Counting labeled unrooted level-1 networks

Unrooted level-1 networks:
explicit formula for *n* leaves, *c* cycles, *m* edges involved in the cycles.

Pointing + bijection:
Bijection between labeled unrooted level-1 networks with *n*+1 leaves and labeled pointed level-1 networks with *n* leaves.

Recursive decomposition of pointed level-1 networks with *n* leaves:



Exponential generating function:

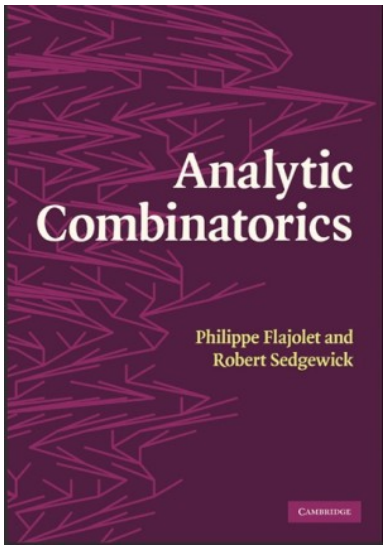
$$G = z + \frac{1}{2}G^2 + \frac{1}{2}\frac{G^2}{(1-G)}$$

Semple & Steel, *TCBB*, 2006

Counting labeled unrooted level-1 networks

Exponential generating function:

$$G = z + \frac{1}{2}G^2 + \frac{1}{2}\frac{G^2}{(1-G)}$$

Using the Singular Inversion Theorem (Theorem VI.6 of 

):

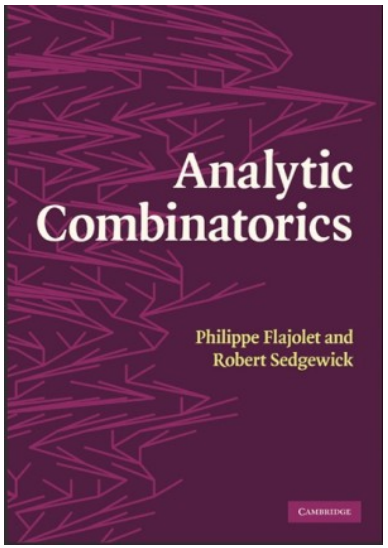
$$g_n \approx 0.2074\,(1.8904)^n\, n^{n-1}$$

Joint work with Mathilde Bouvel and Marefatollah Mansouri

Counting labeled unrooted level-1 networks

Exponential generating function:

$$G = z + \frac{1}{2}G^2 + \frac{1}{2}\frac{G^2}{(1-G)}$$

Using the Singular Inversion Theorem (Theorem VI.6 of 

):

$$g_n \approx 0.2074\ (1.8904)^n\ n^{n-1}$$

Proof :

We write $G = z\,\varphi(G)$, with $\varphi(z) = \dfrac{1}{1 - \frac{1}{2}\,z\,(1+1/(1-z))}$

Then $g_n \approx n!\ \sqrt{\dfrac{\varphi(\tau)}{2\varphi''(\tau)}}\ \dfrac{\rho^{-n}}{\sqrt{\pi\, n^3}}$, with $\rho = \tau / \varphi(\tau)$

and τ is the solution of $\varphi(z)\text{-}z\varphi'(z)=0$

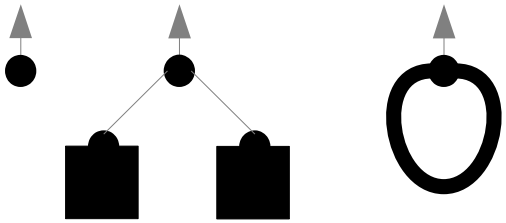
Maplesoft SAGE

Joint work with Mathilde Bouvel and Marefatollah Mansouri

Counting labeled unrooted level-2 networks

Recursive decomposition of pointed level-2 networks with *n* leaves:

z $\frac{1}{2}$ U^2 $\frac{1}{2}$ $\frac{U^2}{1-U}$ *or...*



Seq$_{\geq 1}$, any direction

Seq$_{\geq 2}$, any direction

Seq$_{\geq 1}$

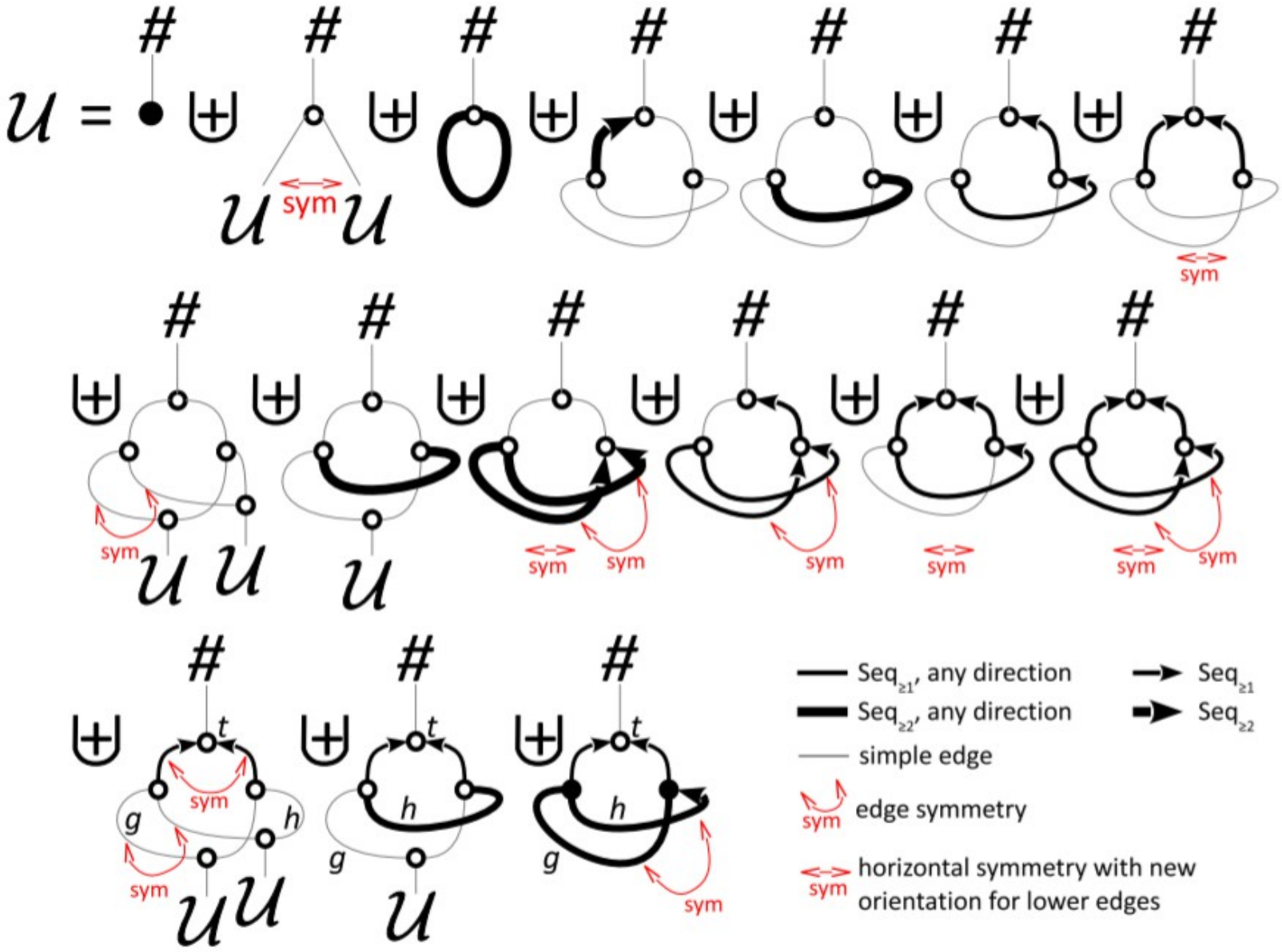
Seq$_{\geq 2}$

simple edge

sym edge symmetry

sym horizontal symmetry with new orientation for lower edges

Counting labeled unrooted level-2 networks


$\mathcal{U}$ =
sym
sym
sym

Counting labeled unrooted level-2 networks

Recursive decomposition of pointed level-2 networks with *n* leaves:

$$\begin{aligned} U &= z + \frac{U^2}{2} + \frac{U^2}{2(1-U)} + \frac{U^2}{1-U} + \frac{U^2}{2(1-U)} + \frac{U^2}{(1-U)^2} + \frac{U^2}{2(1-U)^2} \\ &+ \frac{U^2}{2} + \frac{U^3}{2(1-U)} + \frac{U^4}{4(1-U)^2} + \frac{U^3}{2(1-U)^3} + \frac{U^3}{2(1-U)^2} + \frac{U^4}{2(1-U)^3} \\ &+ \frac{U^4}{4(1-U)^2} + \frac{U^5}{2(1-U)^3} + \frac{U^6}{4(1-U)^4} \end{aligned}$$

Rewrite:

$$U(z) = z\phi(U(z)) \text{ where } \phi(z) = \frac{1}{1 - \frac{3z^5 - 20z^4 + 46z^3 - 46z^2 + 18z}{4(z-1)^4}}.$$

Joint work with Mathilde Bouvel and Marefatollah Mansouri

Counting labeled unrooted level-2 networks

Recursive decomposition of pointed level-2 networks with *n* leaves:

$$\begin{aligned} U &= z + \frac{U^2}{2} + \frac{U^2}{2(1-U)} + \frac{U^2}{1-U} + \frac{U^2}{2(1-U)} + \frac{U^2}{(1-U)^2} + \frac{U^2}{2(1-U)^2} \\ &+ \frac{U^2}{2} + \frac{U^3}{2(1-U)} + \frac{U^4}{4(1-U)^2} + \frac{U^3}{2(1-U)^3} + \frac{U^3}{2(1-U)^2} + \frac{U^4}{2(1-U)^3} \\ &+ \frac{U^4}{4(1-U)^2} + \frac{U^5}{2(1-U)^3} + \frac{U^6}{4(1-U)^4} \end{aligned}$$

Rewrite:

$$U(z) = z\phi(U(z)) \text{ where } \phi(z) = \frac{1}{1 - \frac{3z^5 - 20z^4 + 46z^3 - 46z^2 + 18z}{4(z-1)^4}}.$$

Lagrange inversion:

$$u_n = n![z^n]U(z) = \frac{n!}{n}[z^{n-1}]\phi^n(z)$$

Taylor expansions of $\varphi^n(z)$:

number of leaves	2	3	4	5	6	7
unrooted level-2	-	9	282	14 697	1 071 750	100 467 405

Joint work with Mathilde Bouvel and Marefatollah Mansouri

Counting labeled unrooted level-2 networks

Recursive decomposition of pointed level-2 networks with *n* leaves:

$$\begin{aligned} U &= z + \frac{U^2}{2} + \frac{U^2}{2(1-U)} + \frac{U^2}{1-U} + \frac{U^2}{2(1-U)} + \frac{U^2}{(1-U)^2} + \frac{U^2}{2(1-U)^2} \\ &+ \frac{U^2}{2} + \frac{U^3}{2(1-U)} + \frac{U^4}{4(1-U)^2} + \frac{U^3}{2(1-U)^3} + \frac{U^3}{2(1-U)^2} + \frac{U^4}{2(1-U)^3} \\ &+ \frac{U^4}{4(1-U)^2} + \frac{U^5}{2(1-U)^3} + \frac{U^6}{4(1-U)^4} \end{aligned}$$

Rewrite:

$$U(z) = z\phi(U(z)) \text{ where } \phi(z) = \frac{1}{1 - \frac{3z^5-20z^4+46z^3-46z^2+18z}{4(z-1)^4}}.$$

Lagrange inversion:

$$u_n = n![z^n]U(z) = \frac{n!}{n}[z^{n-1}]\phi^n(z)$$

Taylor expansions + Newton formula

$$u_n = (n-1)! \sum_{\substack{0\le s\le q\le p\le k\le i\le n-1 \\ j=n-1-i-k-p-q-s\ge 0 \\ i\neq 0}} \binom{n+i-1}{i}\binom{4i+j-1}{j}\binom{i}{k}\binom{k}{p}\binom{p}{q}\binom{q}{s} \times \left(\frac{-3}{20}\right)^s \left(\frac{9}{2}\right)^i \left(\frac{-23}{9}\right)^k (-1)^p \left(\frac{-10}{23}\right)^q.$$

Joint work with Mathilde Bouvel and Marefatollah Mansouri

Counting labeled level-*k* networks

Unrooted level-1 networks:
explicit formula for *n* leaves, *c* cycles, *m* edges involved in the cycles

Semple & Steel, *TCBB*, 2006

+ asymptotic evaluation for *n* leaves: $\approx 0.207\ (1.890)^n\ n^{n-1}$

Rooted level-1 networks :
Explicit formula for *n* leaves, *c* cycles, *m* edges across cycles
+ asymptotic evaluation for *n* leaves: $\approx 0.134\ (2.943)^n\ n^{n-1}$

Unrooted level-2 networks :
Explicit formula for *n* leaves :

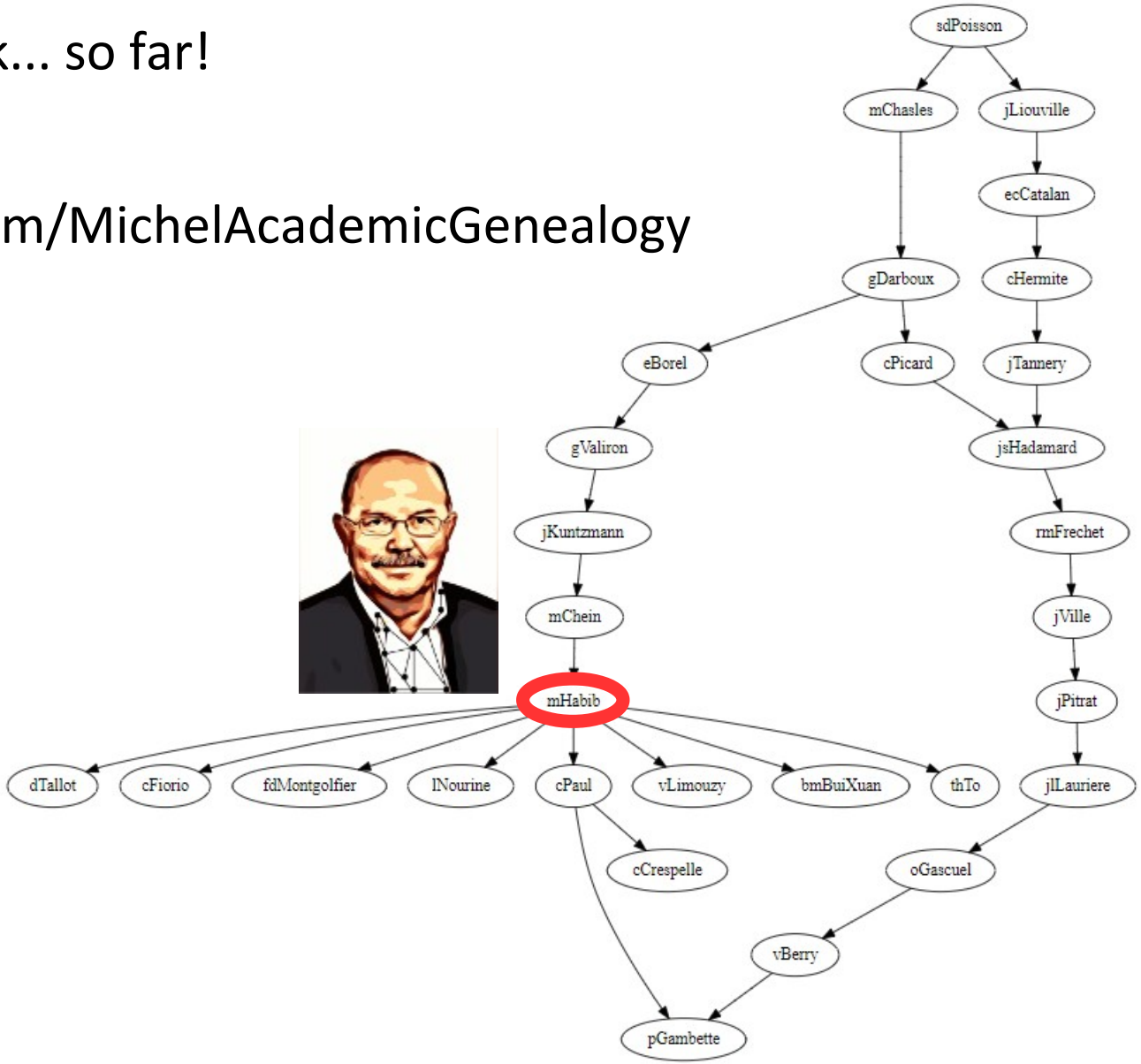
$$(n\text{-}1)! \sum_{\substack{0\le s\le q\le p\le k\le i\le n-1 \\ j=n-1-i-k-p-q-s\ge 0 \\ i\neq 0}} \binom{n+i-1}{i}\binom{4i+j-1}{j}\binom{i}{k}\binom{k}{p}\binom{p}{q}\binom{q}{s}\left(\frac{-3}{20}\right)^s\left(\frac{9}{2}\right)^i\left(\frac{-23}{9}\right)^k(-1)^p\left(\frac{-10}{23}\right)^q$$

number of leaves	2	3	4	5	6	7
unrooted level-1	-	2	15	192	3 450	79 740
rooted level-1	3	36	723	20 280	730 755	32 171 580
unrooted level-2	-	9	282	14 697	1 071 750	100 467 405

Joint work with Mathilde Bouvel and Marefatollah Mansouri

A message from Mathilde...

The academic phylogenetic network of Michel & friends

A level-2 network... so far!

Add yourself to
https://tinyurl.com/MichelAcademicGenealogy