Phylogenetics Workshop,
Isaac Newton Institute for Mathematical Sciences
Cambridge – 21/06/2011

Quartets and unrooted level-k networks

Philippe Gambette

Outline

- Abstract and explicit phylogenetic networks
- Level-*k* networks
- Unrooted level-1 networks and circular split systems
- Reconstruction from triplets and quartets

Outline

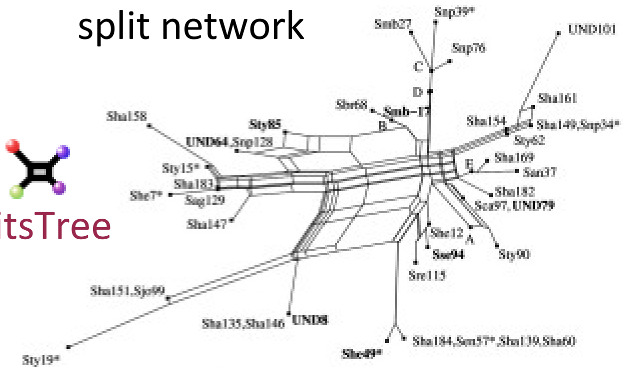
- Abstract and explicit phylogenetic networks
- Level-*k* networks
- Unrooted level-1 networks and circular split systems
- Reconstruction from triplets and quartets

Abstract and explicit phylogenetic networks

Abstract phylogenetic networks:
Visualize reticulate evolution data

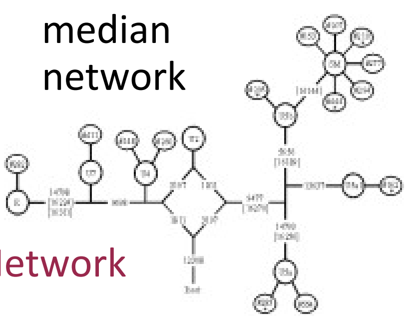
Explicit phylogenetic networks:
Each node can be interpreted as a current or ancestral species

split network



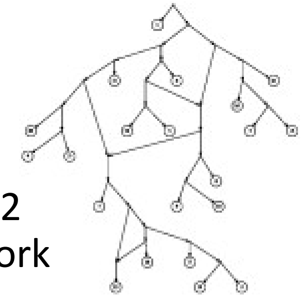
SplitsTree

median network



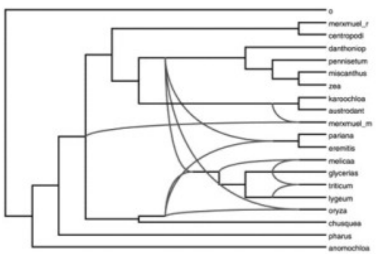
Network

level-2 network



Simplistic

galled network

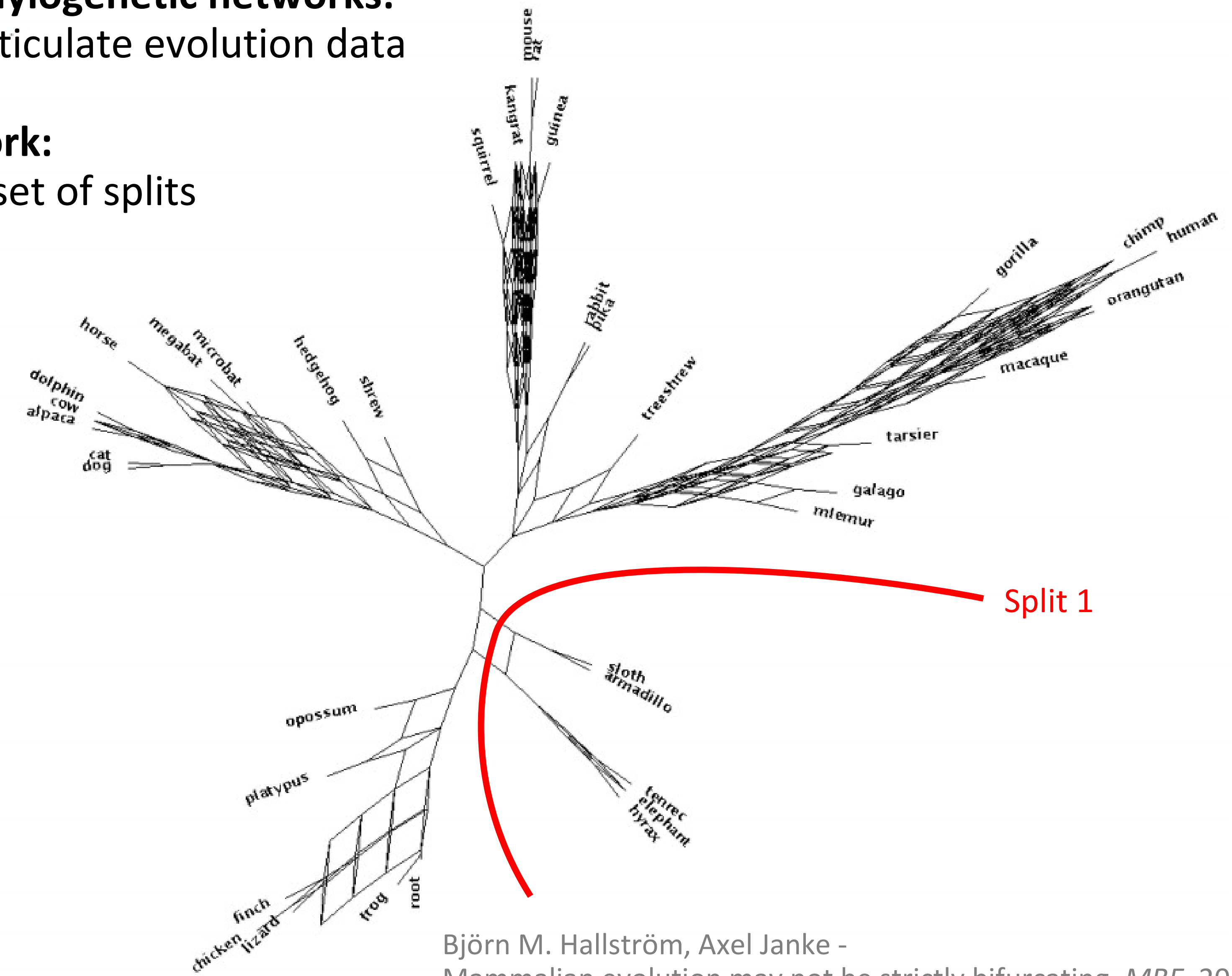


Dendroscope

Split networks

Abstract phylogenetic networks:
Visualize reticulate evolution data

Split network:
Visualize a set of splits

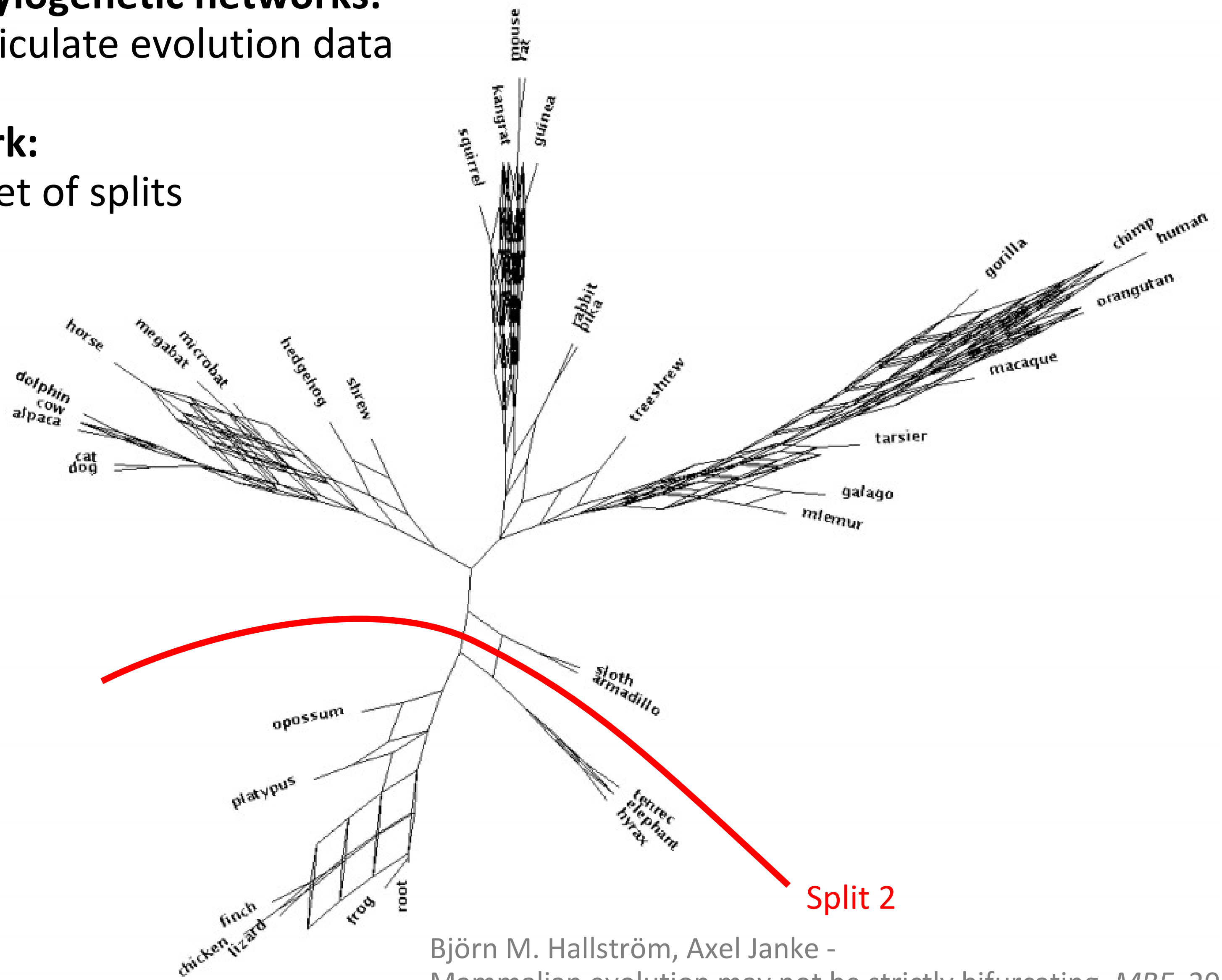


Björn M. Hallström, Axel Janke -
Mammalian evolution may not be strictly bifurcating, *MBE*, 2010

Split networks

Abstract phylogenetic networks:
Visualize reticulate evolution data

Split network:
Visualize a set of splits

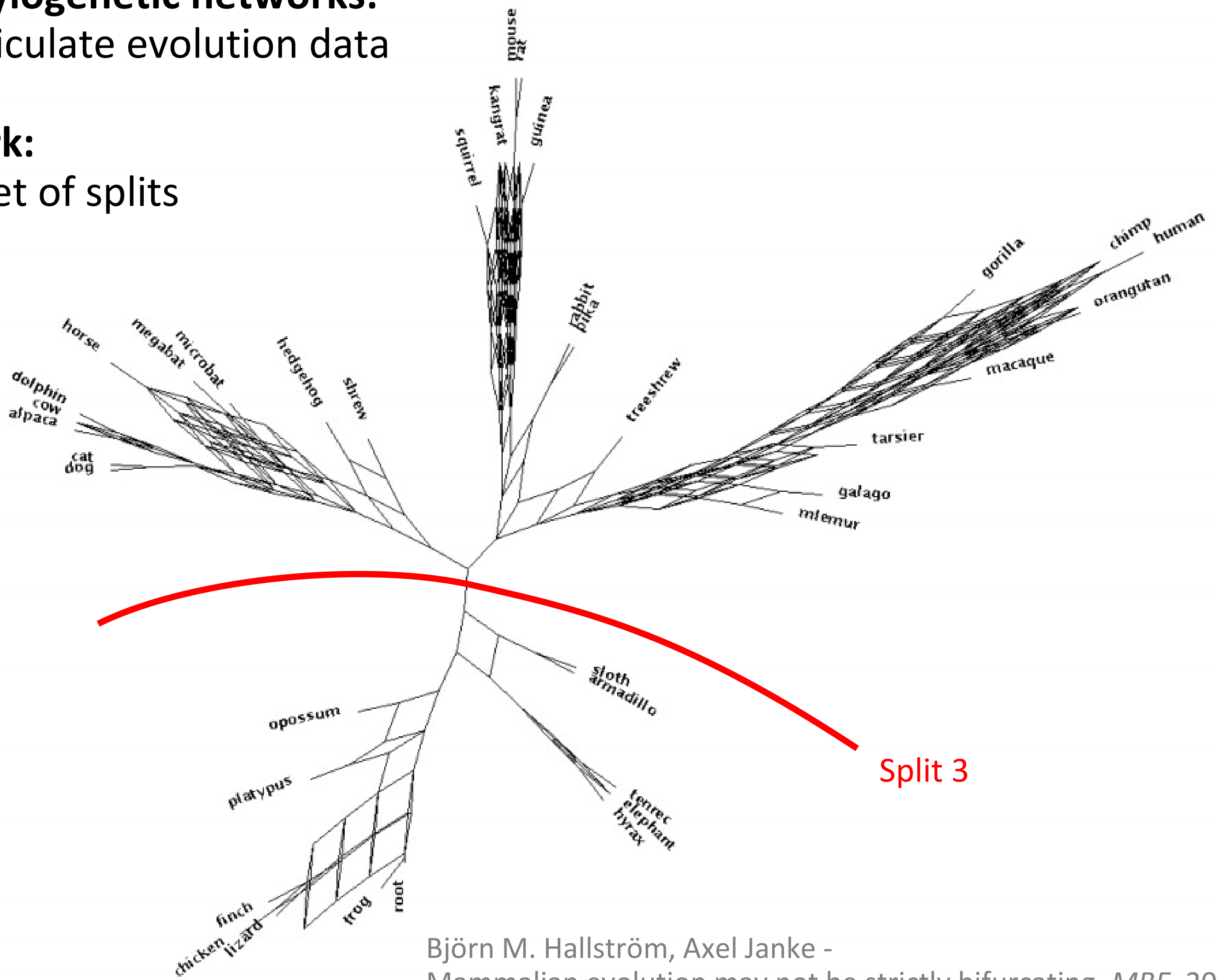


Björn M. Hallström, Axel Janke -
Mammalian evolution may not be strictly bifurcating, *MBE*, 2010

Split networks

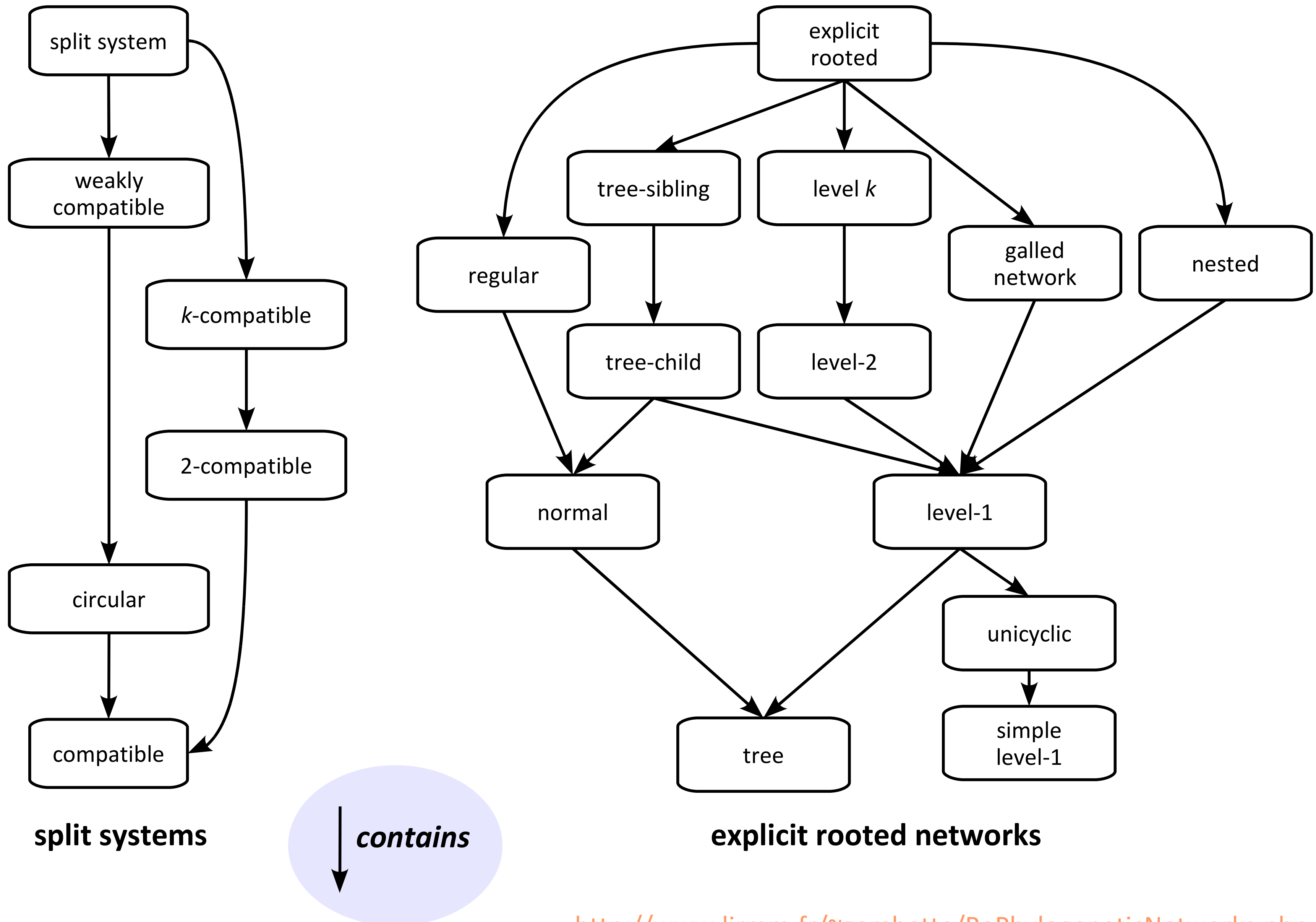
Abstract phylogenetic networks:
Visualize reticulate evolution data

Split network:
Visualize a set of splits



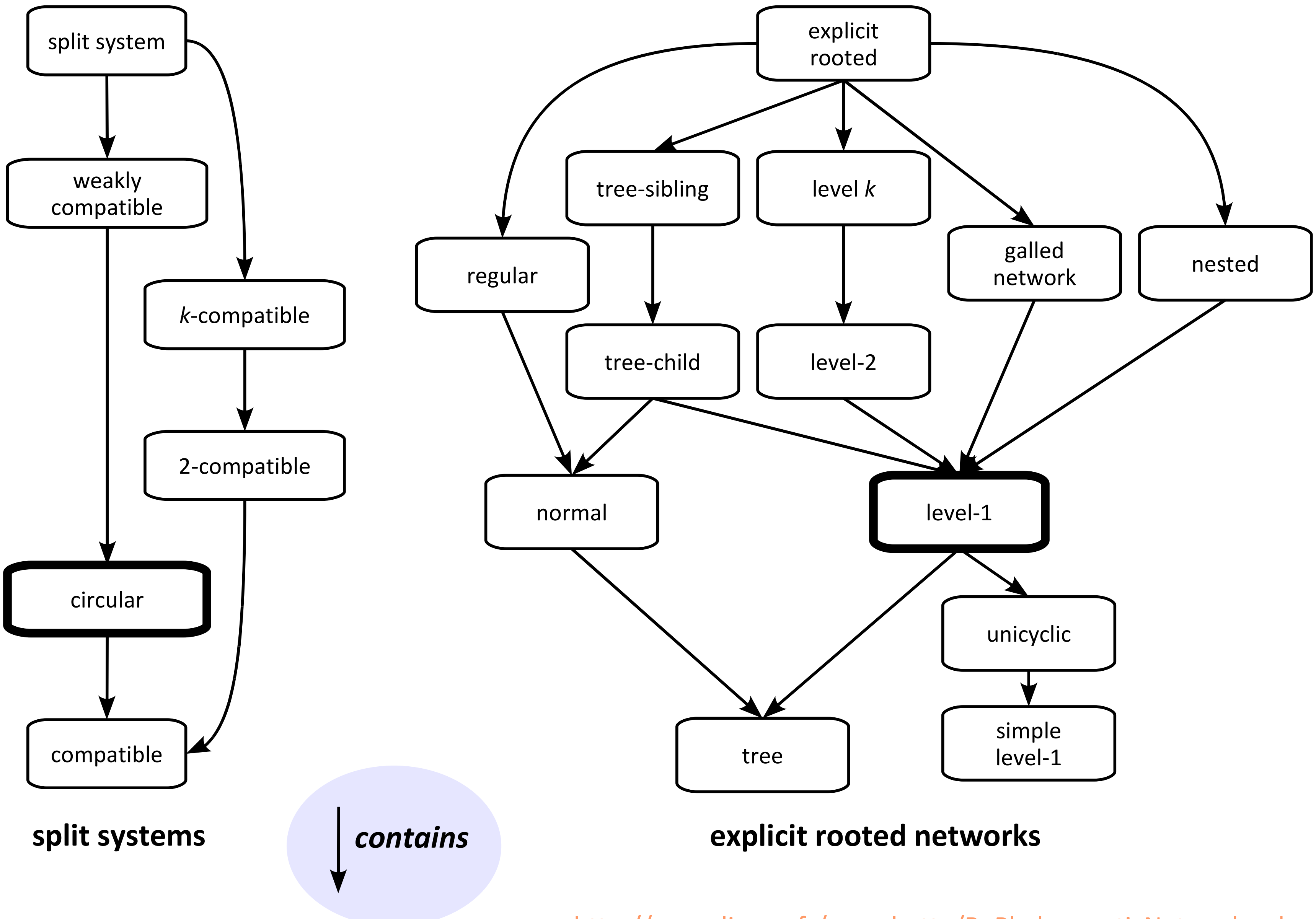
Björn M. Hallström, Axel Janke -
Mammalian evolution may not be strictly bifurcating, *MBE*, 2010

Network subclass hierarchy



http://www.lirmm.fr/~gambette/RePhylogeneticNetworks.php

Network subclass hierarchy



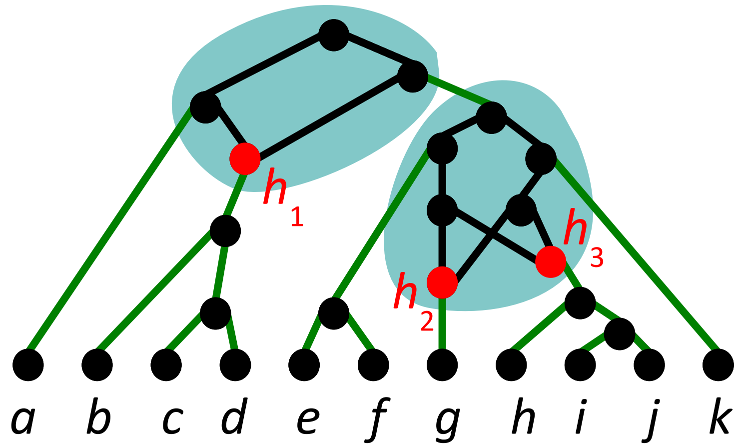
http://www.lirmm.fr/~gambette/RePhylogeneticNetworks.php

Outline

- Abstract and explicit phylogenetic networks
- Level-*k* networks
- Unrooted level-1 networks and circular split systems
- Reconstruction from triplets and quartets

Level-*k* networks

level: how “far” is the network from a tree ?
small level ➡ tree structure ➡ fast algorithms



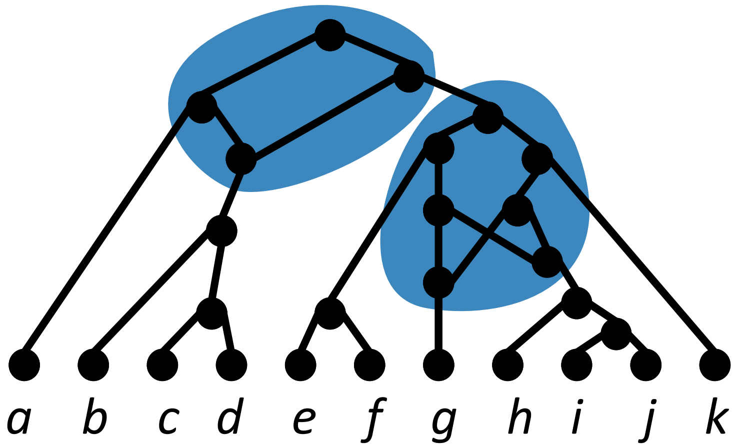
level-2 network

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

Choy, Jansson, Sadakane & Sung, *TCS*, 2005

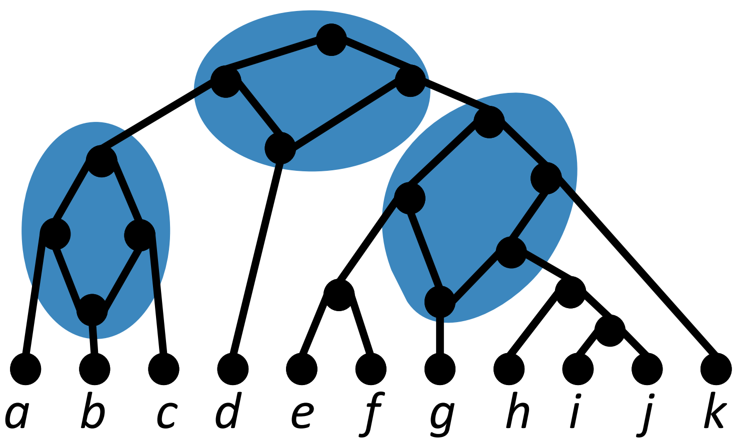
Level-*k* networks

level: how “far” is the network from a tree ?
small level ➡ tree structure ➡ fast algorithms



level-2 network

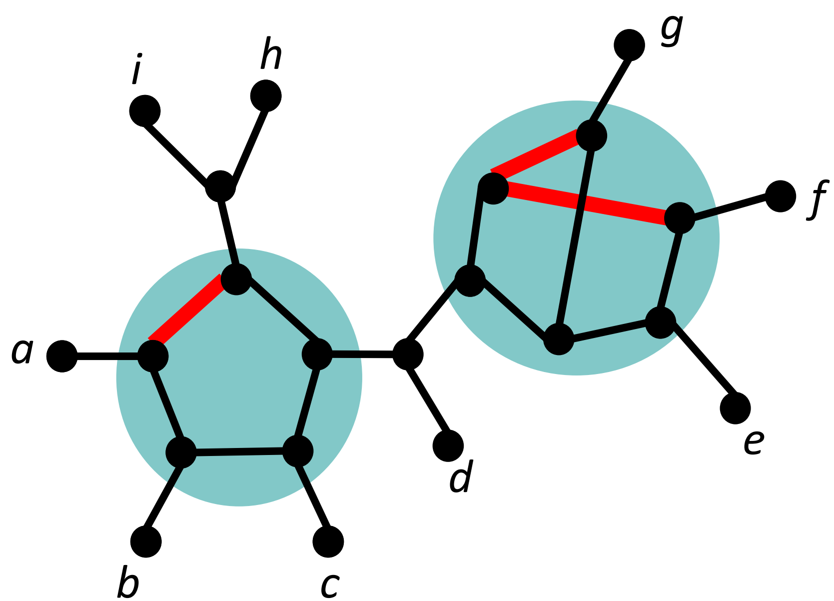
level =
maximum number of reticulations
by *blob*.



level-1 network
(“*galled tree*”)

Unrooted level-*k* networks

level: how “far” is the network from an unrooted tree ?
small level ➡ tree structure ➡ fast algorithms

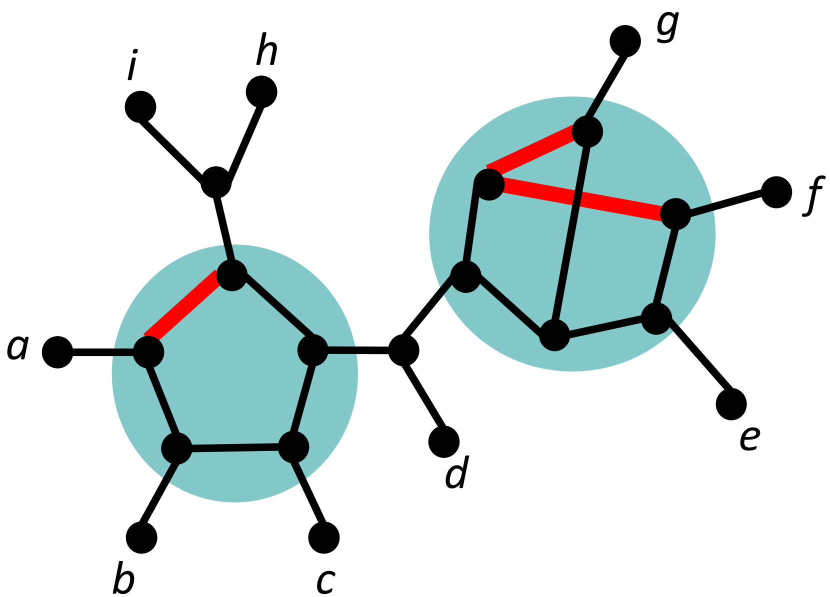


unrooted level-2 network

level =
maximum number of edges to remove, by *blob*, to obtain a tree.

Unrooted level-*k* networks

level: how "far" is the network from an unrooted tree ?
small level ➡ tree structure ➡ fast algorithms

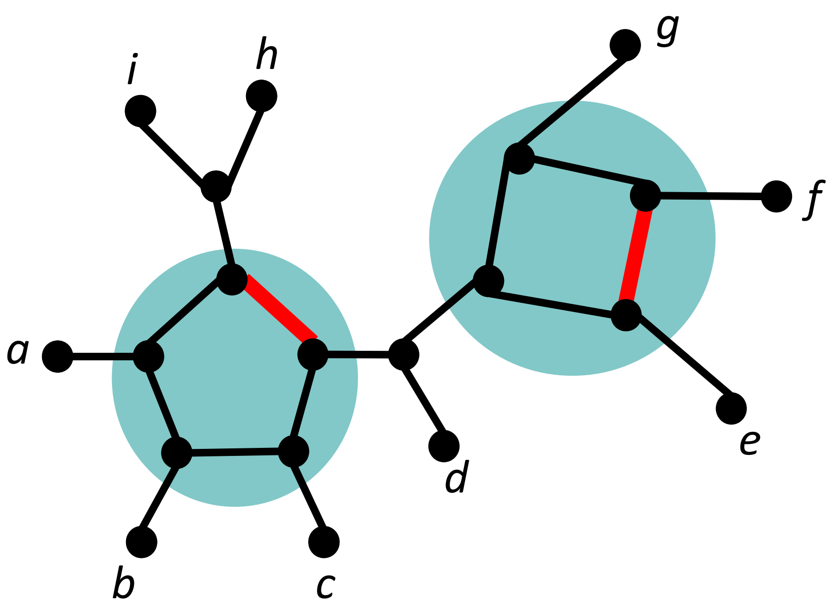


unrooted level-2 network

level =
maximum number of edges to remove, by *blob*, to obtain a tree.
= maximum ***cyclomatic number*** of the blobs

Unrooted level-*k* networks

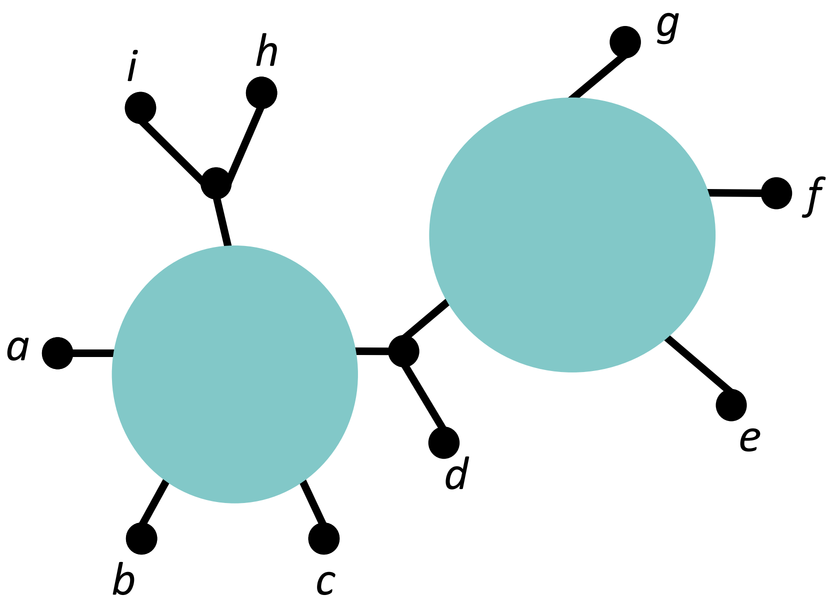
level: how “far” is the network from an unrooted tree ?
small level ➡ tree structure ➡ fast algorithms



level =
maximum number of edges to remove, by *blob*, to obtain a tree.

unrooted level-1 network ➡ tree of cycles
(unrooted galled tree)

Unrooted level-*k* networks

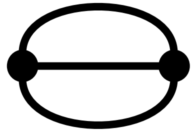


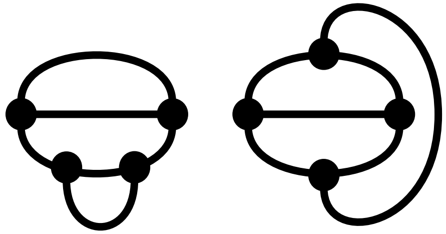
level =
maximum number of edges to remove, by *blob*, to obtain a tree.

unrooted level-*k* network ➡ tree of blobs
➡ tree of **generators** of level ≤*k*

Unrooted level-*k* generators: bridgeless loopless 3-regular multigraphs with 2*k*-2 vertices

level-2 generator





level-3 generators

Gambette, Berry & Paul, manuscript, 2011

Counting labeled level-*k* networks

Unrooted level-1 networks:
explicit formula for *n* leaves, *c* cycles, *m* edges across cycles Semple & Steel, *TCBB*, 2006

Counting labeled level-*k* networks

Unrooted level-1 networks:
explicit formula for n leaves, c cycles, m edges across cycles — Semple & Steel, *TCBB*, 2006

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

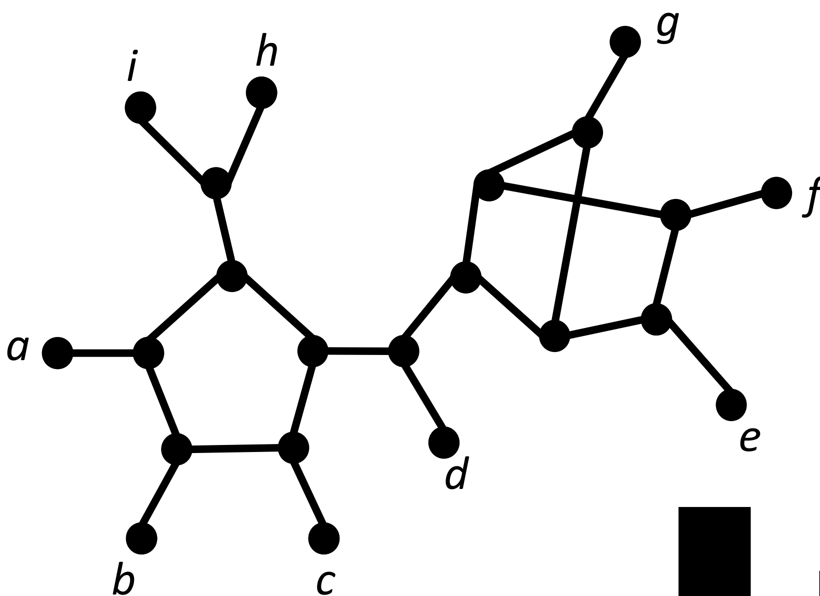
Rooted level-1 networks :
Explicit formula for n leaves, c cycles, m edges across cycles
+ asymptotic evaluation for n leaves: $\approx 0.134\ 2.943^n\ n^{n-1}$

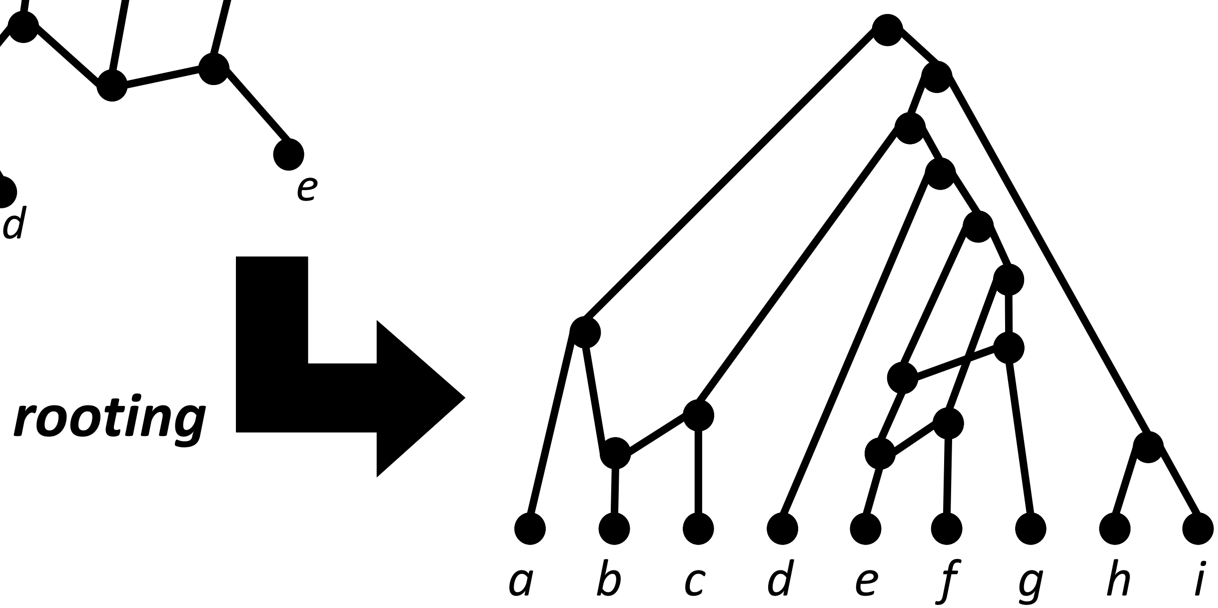
Unrooted level-2 networks :
Explicit formula for n leaves : $$(n-1)! \sum_{\substack{0 \le s \le q \le p \le k \le i \le n-1 \\ j=n-1-i-k-p-q-s \ge 0 \\ i \ne 0}} \binom{n+i-1}{i} \binom{4i+j-1}{j} \binom{i}{k} \binom{k}{p} \binom{p}{q} \binom{q}{s} \left(\frac{-4}{21}\right)^s \left(\frac{9}{2}\right)^i \left(\frac{23}{9}\right)^k (-1)^p \left(\frac{-21}{46}\right)^q$$

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	1 071 720	100 461 195

Bouvel, Gambette, Berry & Paul, manuscript, 2011

Equivalence between rooted and unrooted level

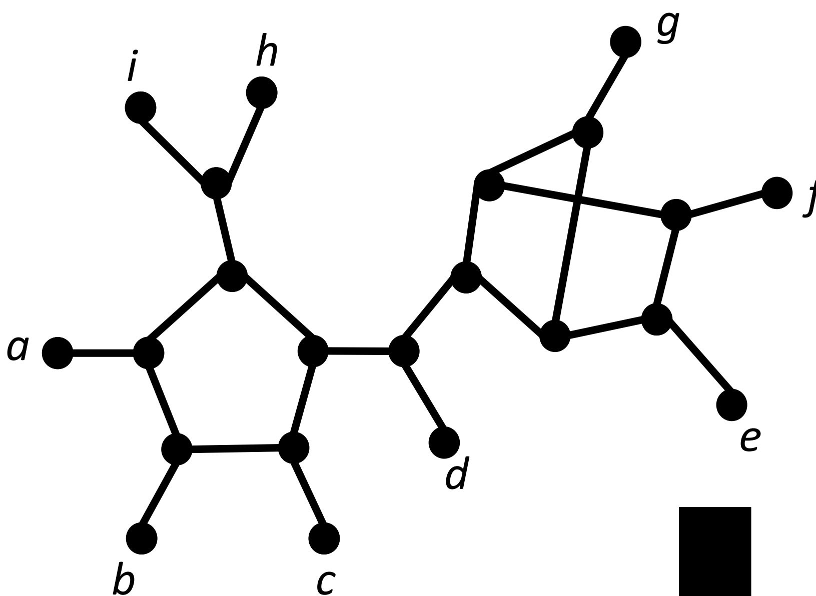


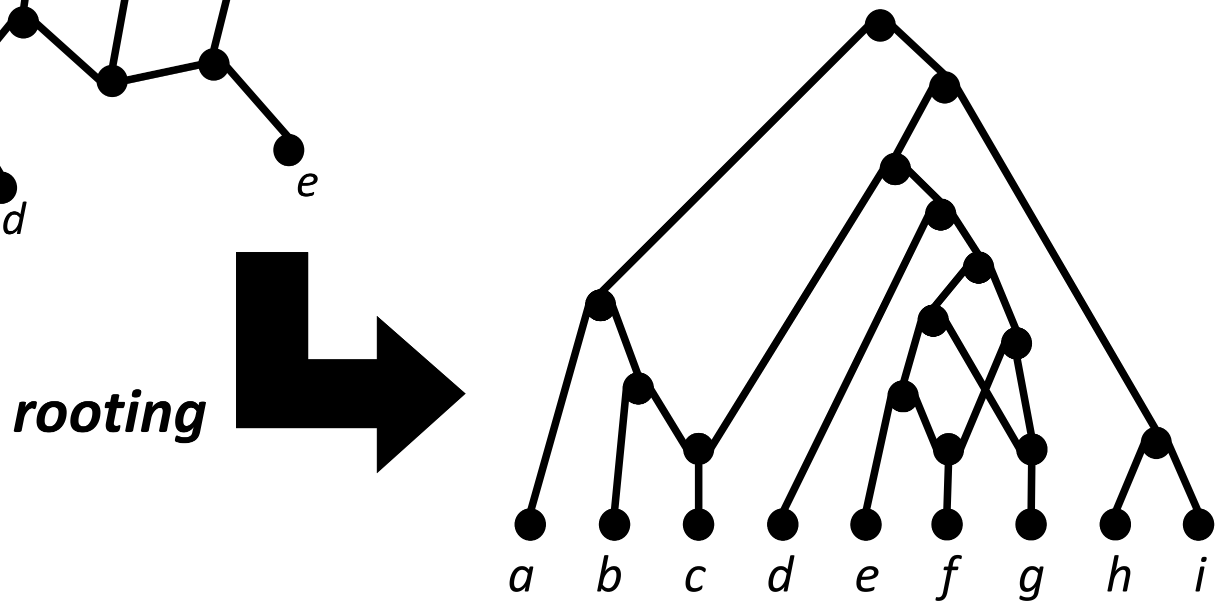


Rooting:
- choosing a root
- choosing an orientation for the edges

Gambette, Berry & Paul, manuscript, 2011

Equivalence between rooted and unrooted level

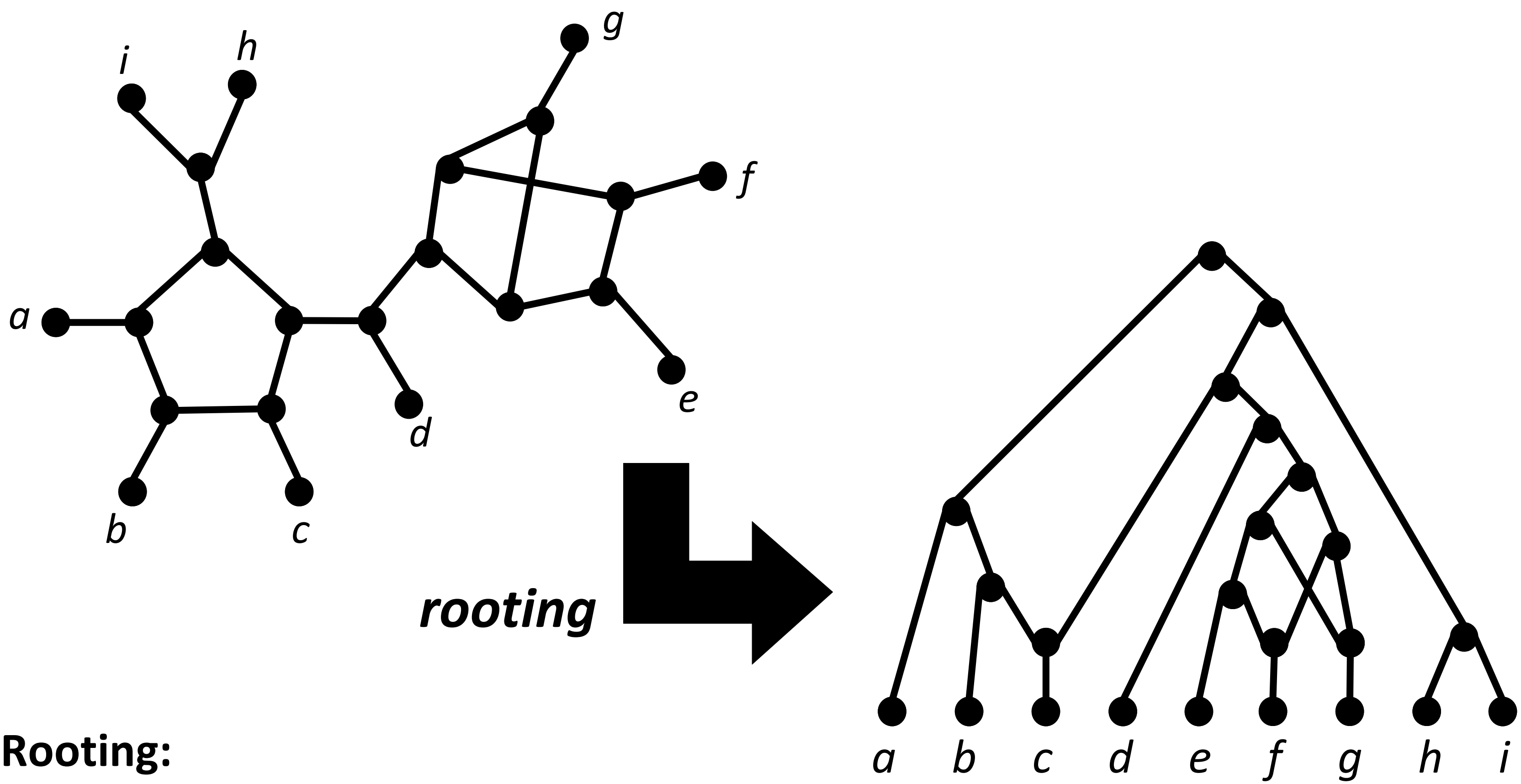




Rooting:
- choosing a root
- choosing an orientation for the edges

Gambette, Berry & Paul, manuscript, 2011

Equivalence between rooted and unrooted level



Rooting:
- choosing a root
- choosing an orientation for the edges

- **many** possible rootings (possibly exponential in the level)
- **same level** (invariant)

Gambette, Berry & Paul, manuscript, 2011

Outline

- Abstract and explicit phylogenetic networks
- Level-*k* networks
- Unrooted level-1 networks and circular split systems
- Reconstruction from triplets and quartets

Splits in unrooted level-*k* networks

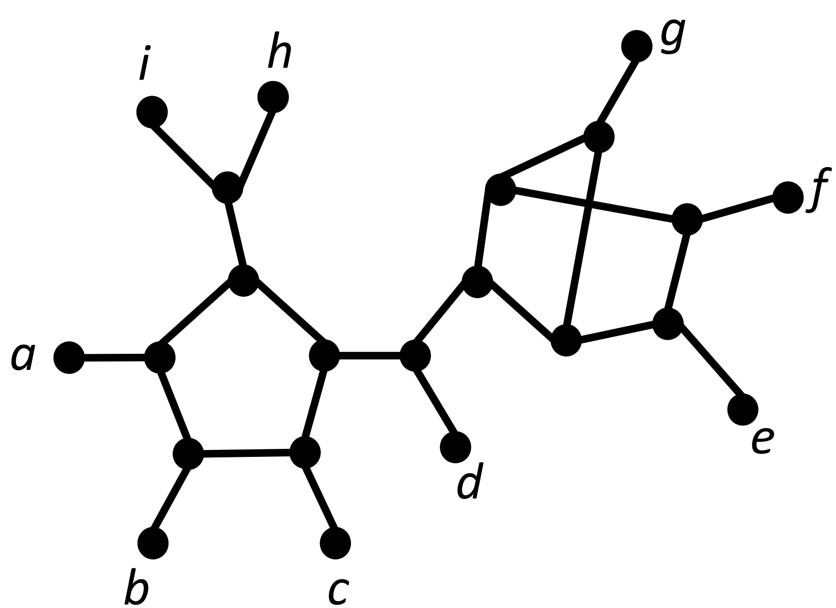
Split:

➡ Split of a tree contained in the network?

Woolley, Posada & Crandall, *PLoS One*, 2008

➡ Leaves separated by a minimal cut in the network?

≈ Brandes & Cornelsen, *DAM*, 2009

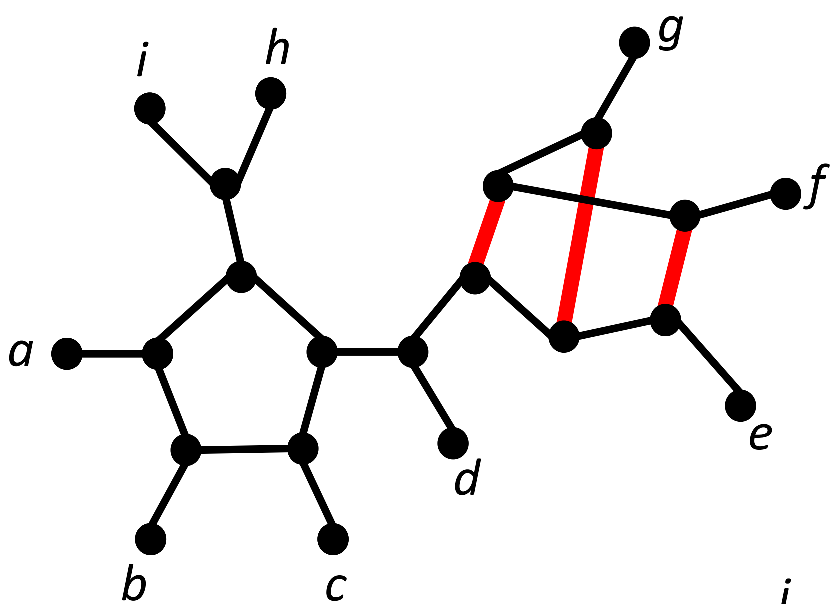


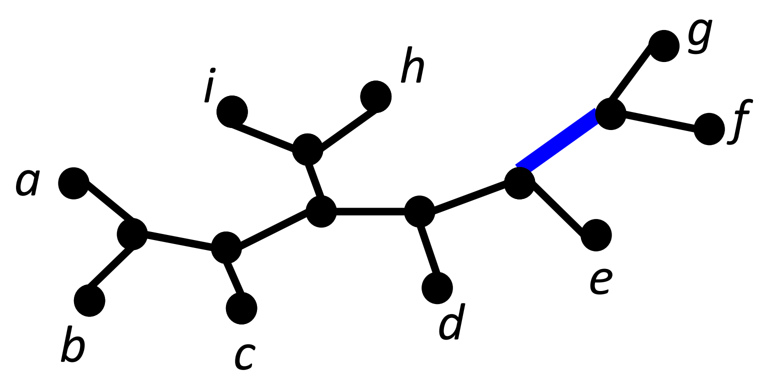
Splits in unrooted level-*k* networks

Bipartition :

- **Split** of a tree contained in the network
- Leaves separated by a **minimal cut** in the network

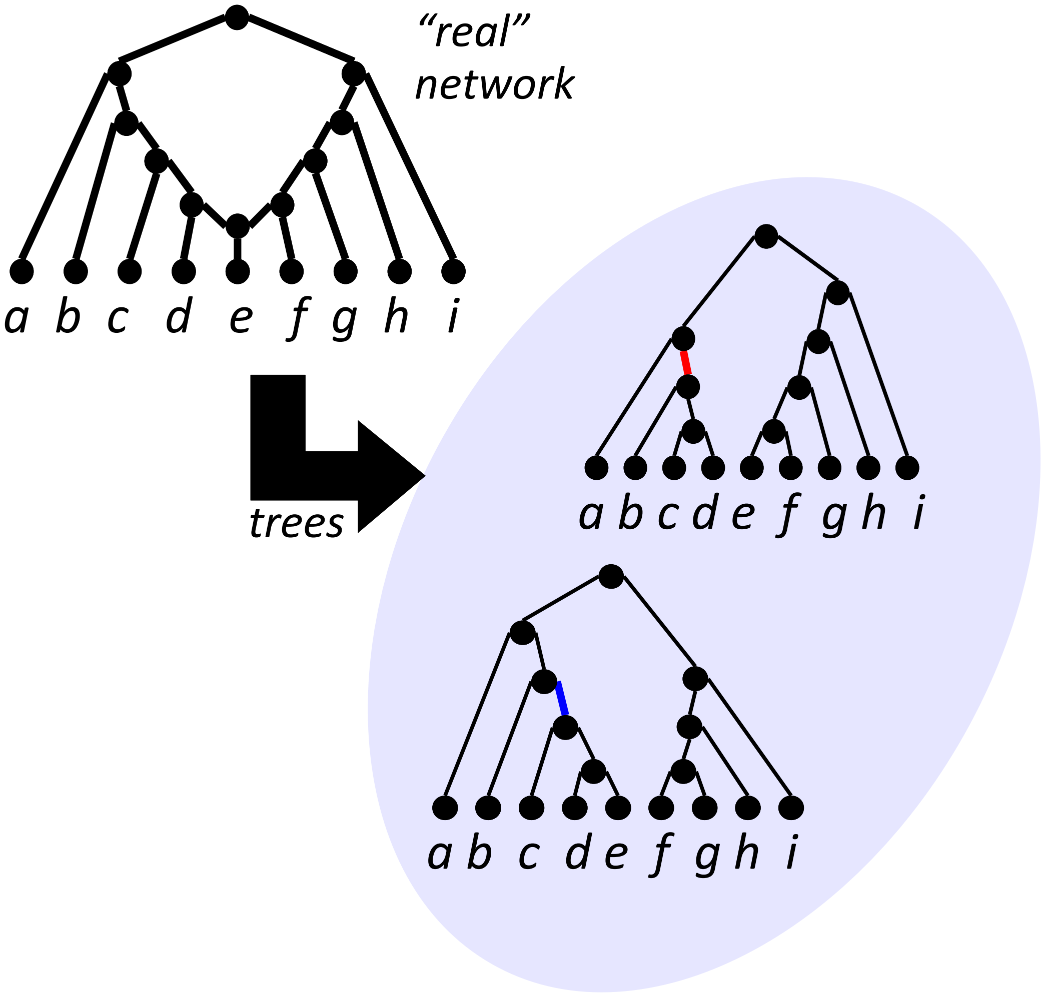
equivalent !



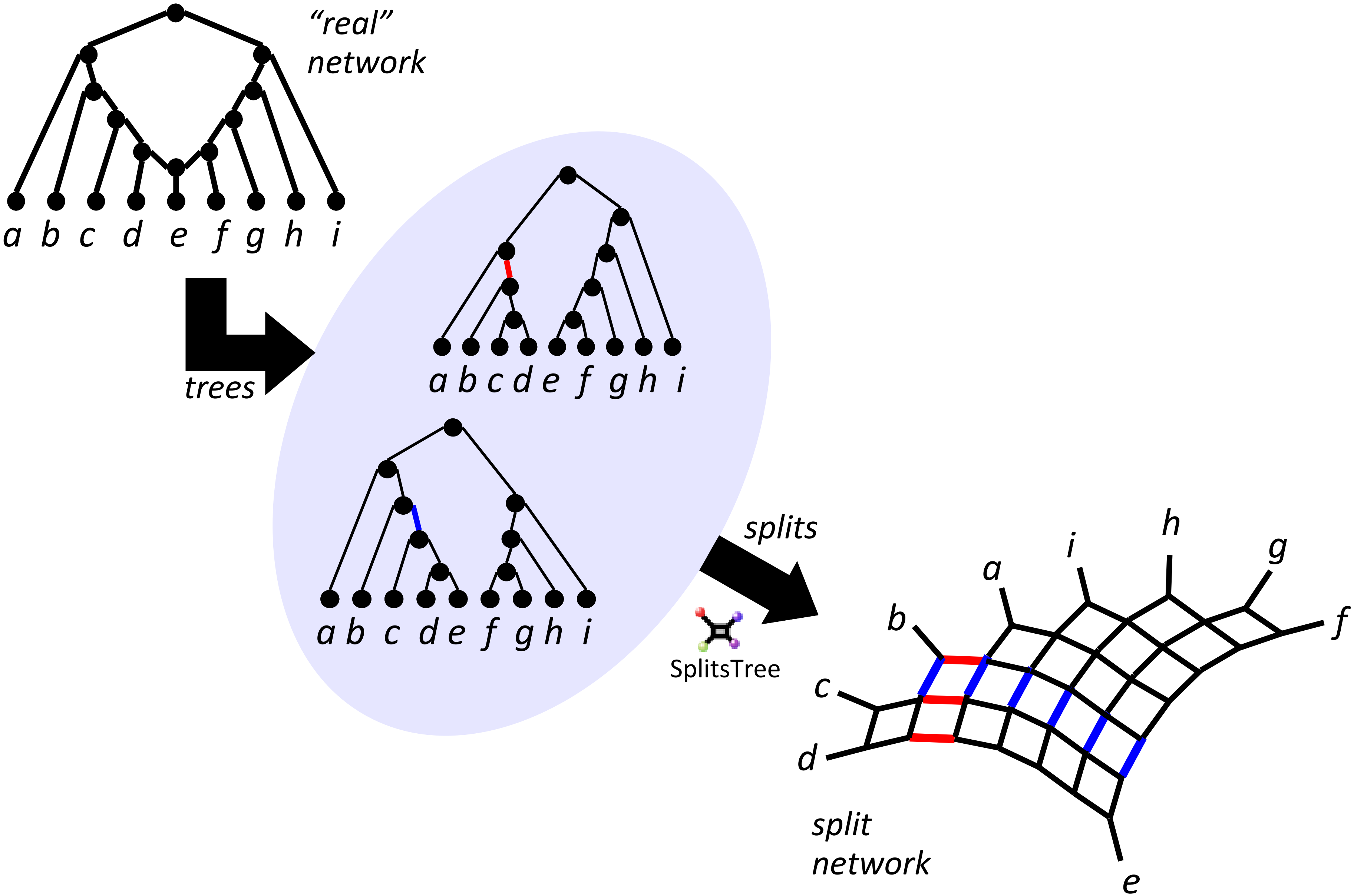


Gambette, Berry & Paul, manuscript, 2011

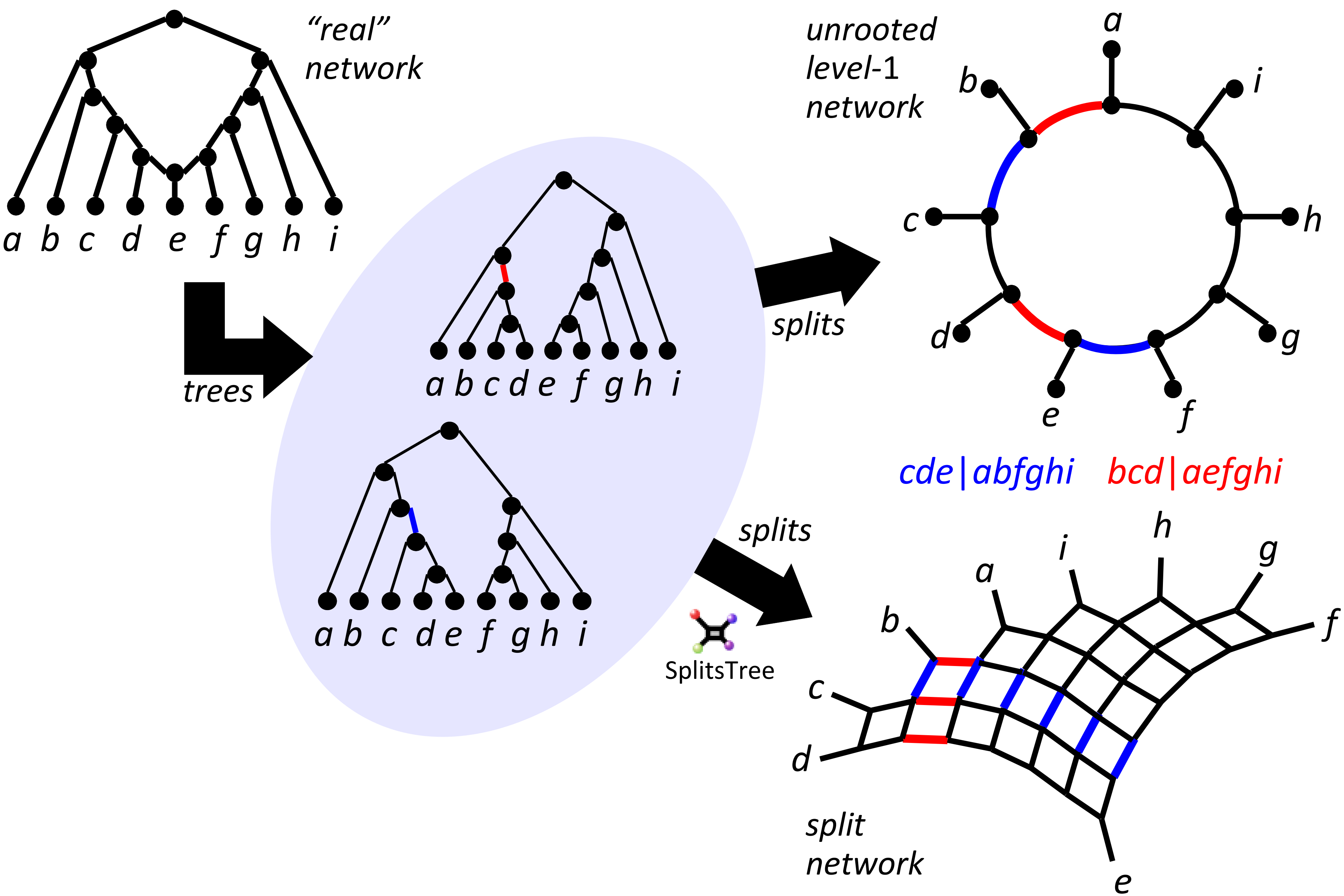
Representing splits by unrooted networks



Representing splits by unrooted networks

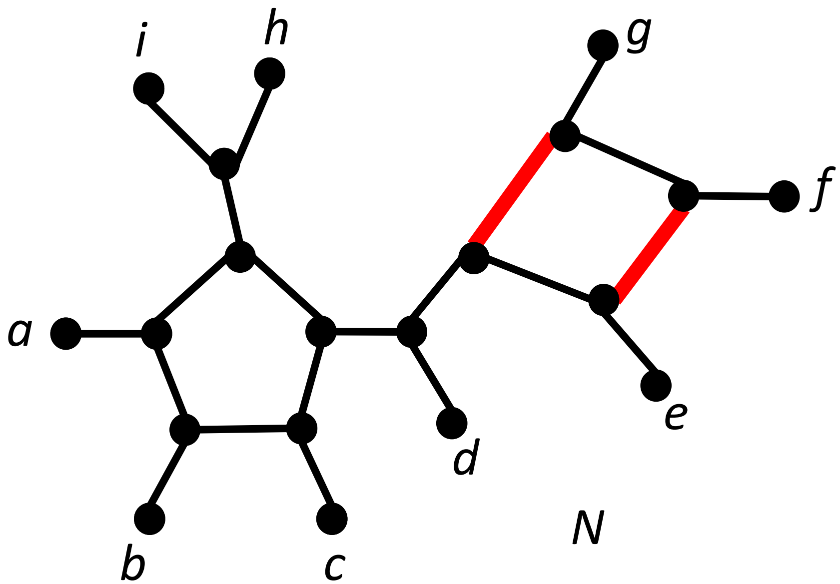


Representing splits by unrooted networks

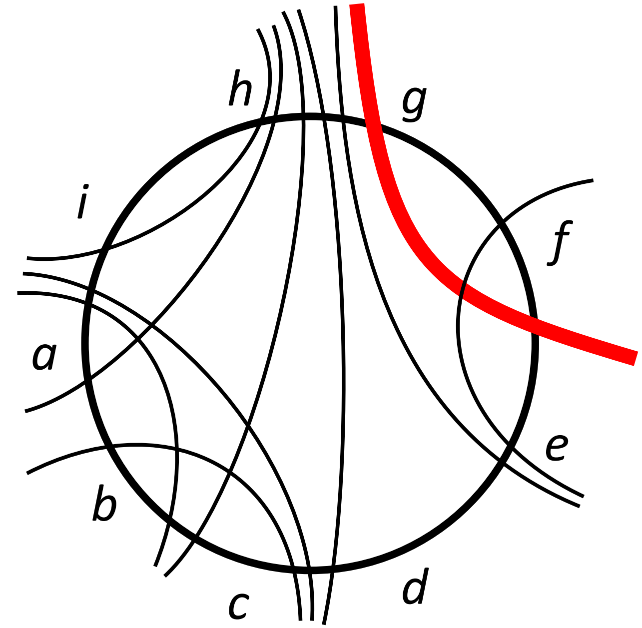


Splits in unrooted level-1 networks

Splits $\Sigma(N)$ of a level-1 network

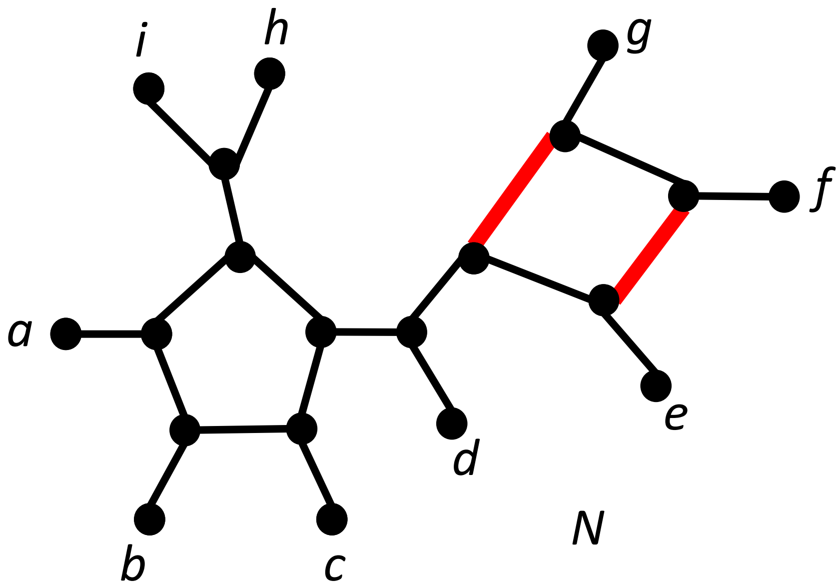


Circular split system Σ

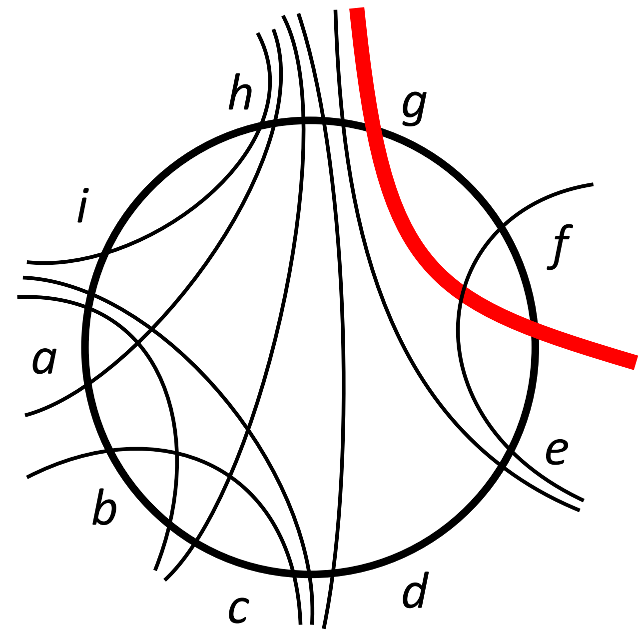


Splits in unrooted level-1 networks

Splits $\Sigma(N)$ of a level-1 network ➡ $\Sigma(N)$ circular



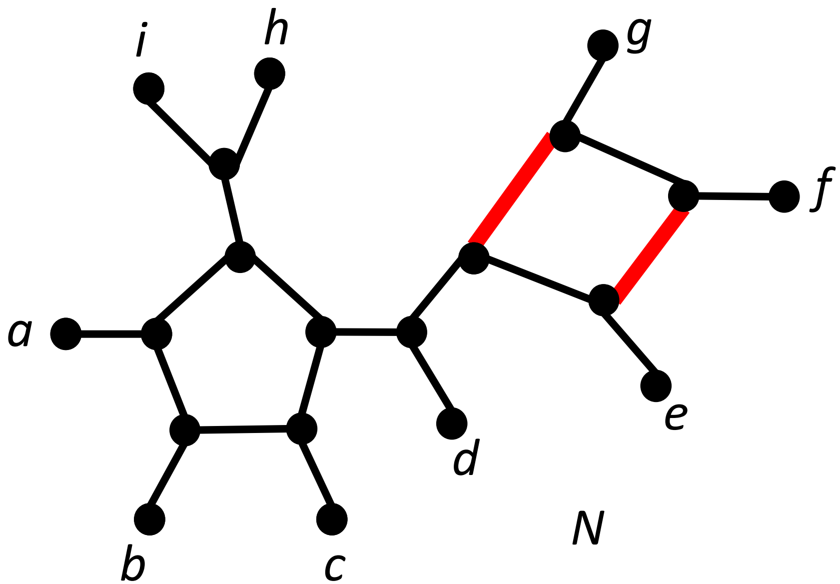
Circular split system Σ



Gambette, Berry & Paul, manuscript, 2011

Splits in unrooted level-1 networks

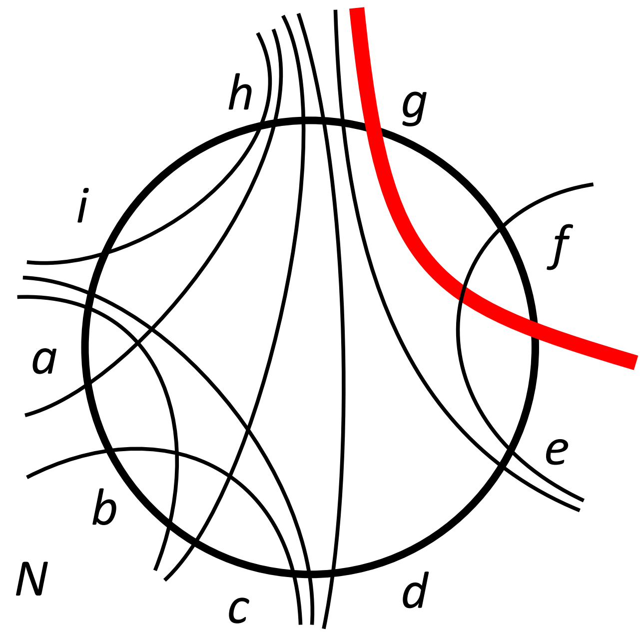
Splits $\Sigma(N)$ of a level-1 network ➡ $\Sigma(N)$ circular



Circular split system Σ

⬇

There exists a level-1 network N such that $\Sigma \subseteq \Sigma(N)$



Gambette, Berry & Paul, manuscript, 2011

Outline

- Abstract and explicit phylogenetic networks
- Level-*k* networks
- Unrooted level-1 networks and circular split systems
- Reconstruction from triplets and quartets

Combinatorial phylogenetic network reconstruction

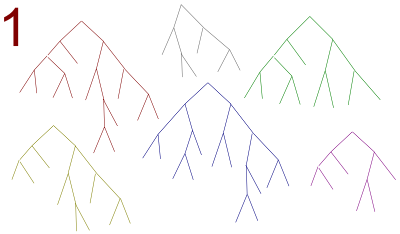
```
species 1 : AATTGCAG TAGCCCAAAAT
species 2 : ACCTGCAG TAGACCAAT
species 3 : GCTTGCCG TAGACAAGAAT
species 4 : ATTTGCAG AAGACCAAAT
species 5 :          TAGACAAGAAT
species 6 : ACTTGCAG TAGCACAAAAT
species 7 : ACCTGGTG TAAAAT
            G1       G2
```

{gene sequences}

↓

{trees}

T1



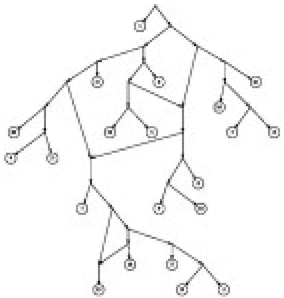
T2

HOGENOM database
Dufayard, Duret, Penel, Gouy, Rechenmann & Perrière, *BioInf*, 2005

Phyl-ARIANE

ANR

↓



network

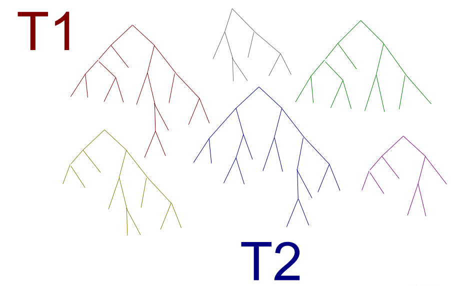
contains the trees
+ "optimal"

Combinatorial phylogenetic network reconstruction

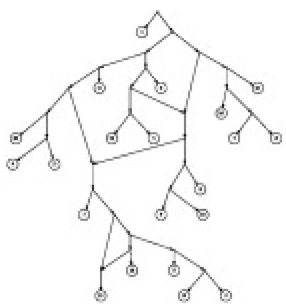
```
species 1 : AATTGCAG TAGCCCAAAAT
species 2 : ACCTGCAG TAGACCAAT
species 3 : GCTTGCCG TAGACAAGAAT
species 4 : ATTTGCAG AAGACCAAAT
species 5 :          TAGACAAGAAT
species 6 : ACTTGCAG TAGCACAAAAT
species 7 : ACCTGGTG TAAAAT
            G1       G2
```

{gene sequences}

{trees}



HOGENOM database
Dufayard, Duret, Penel, Gouy, Rechenmann & Perrière, *BioInf*, 2005
> 500 species, >70 000 trees

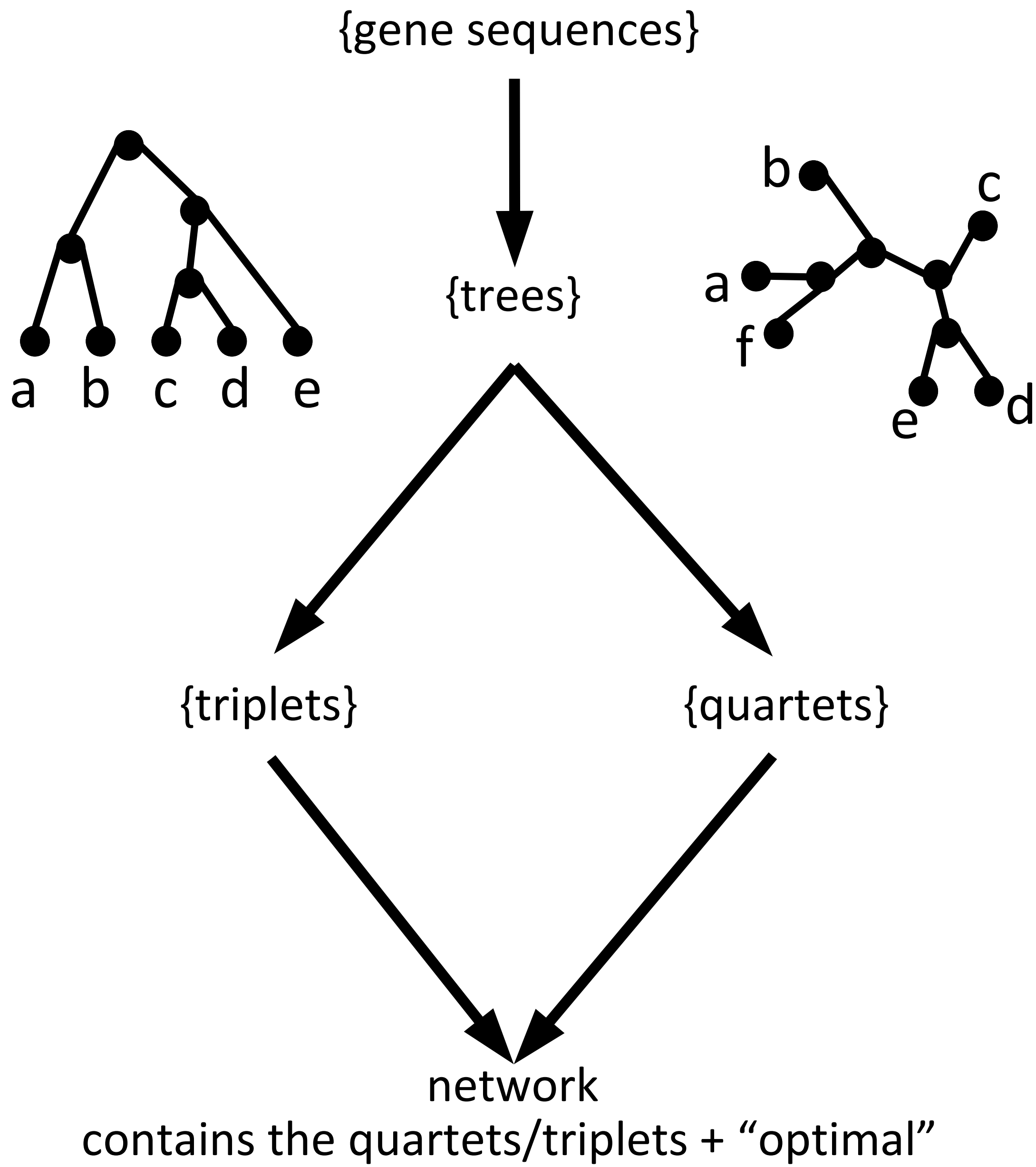


network

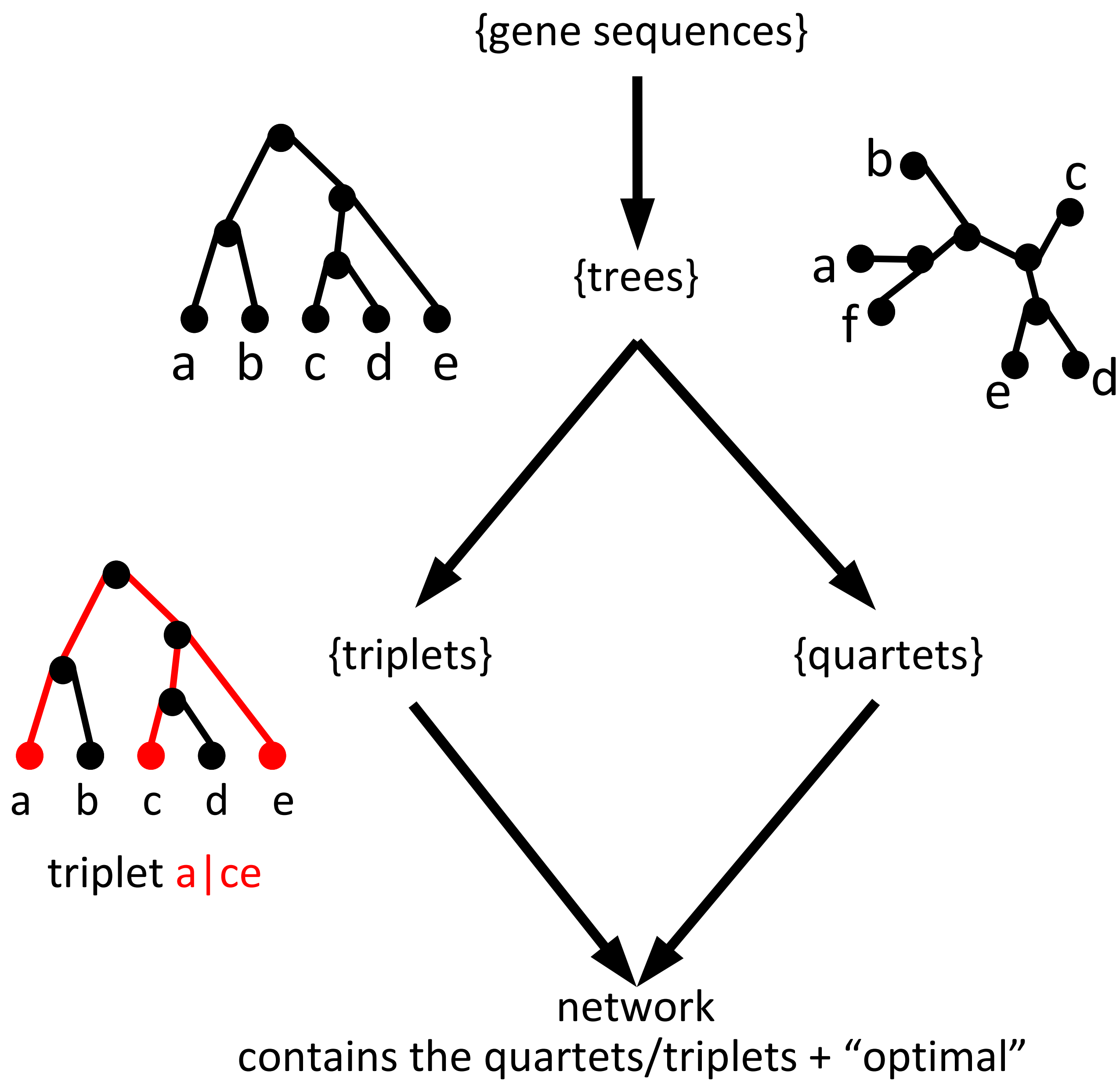
contains the trees
+ “optimal”
NP-complete for 2 rooted trees

Bordewich & Semple, *DAM*, 2007

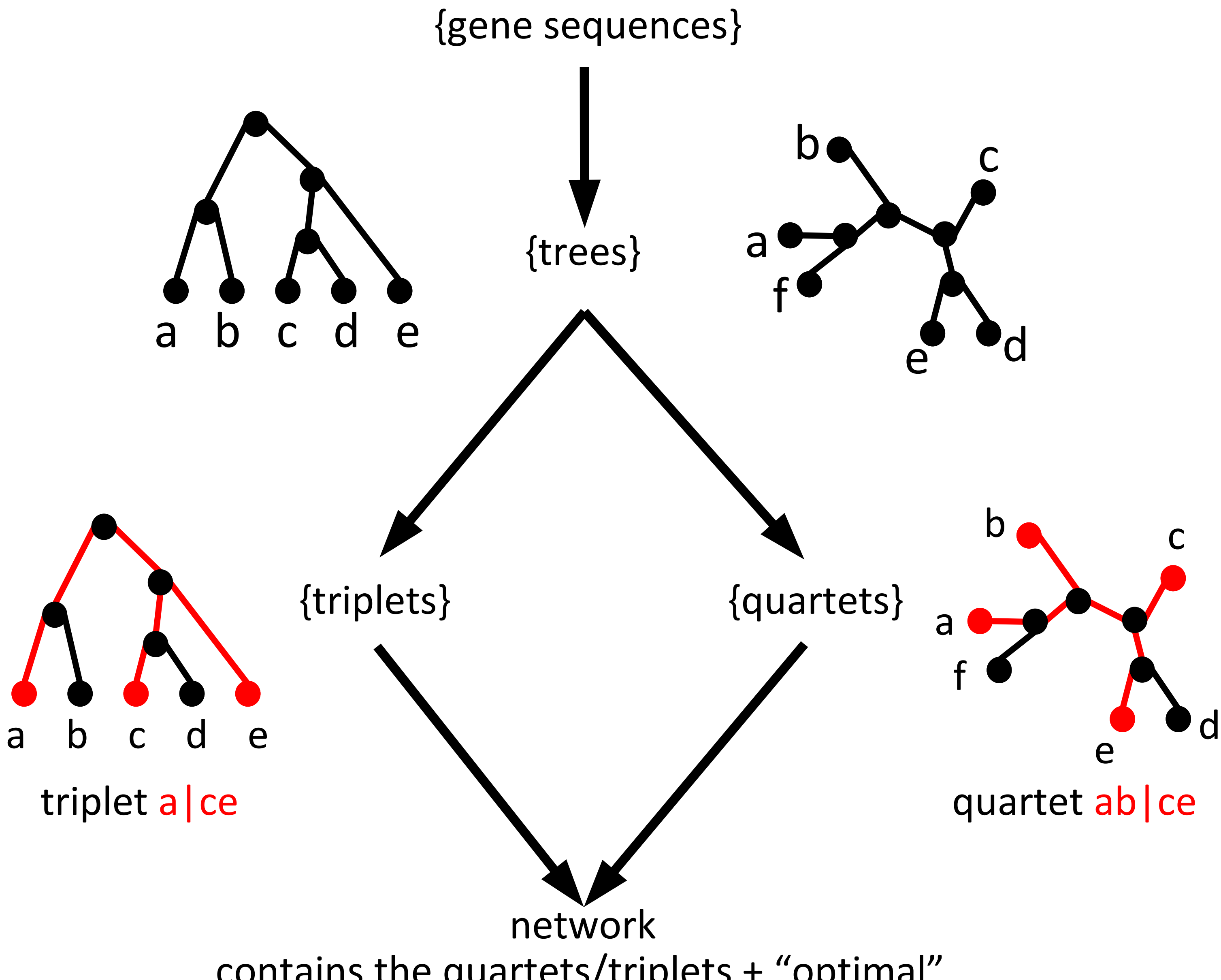
Reconstruction from triplets / quartets



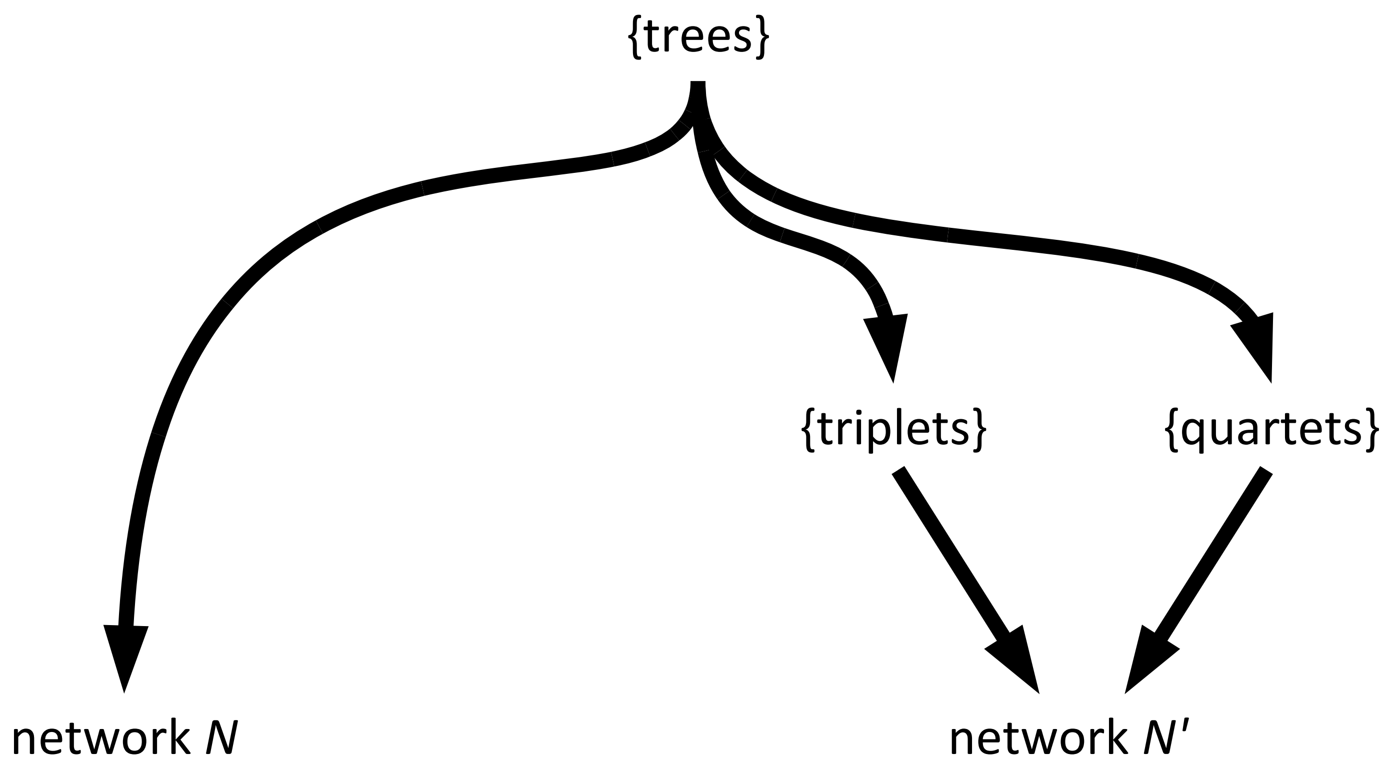
Reconstruction from triplets / quartets



Reconstruction from triplets / quartets



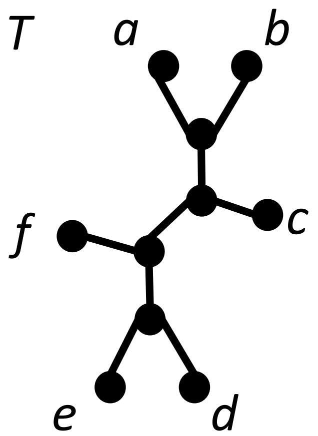
Reconstruction from triplets / quartets

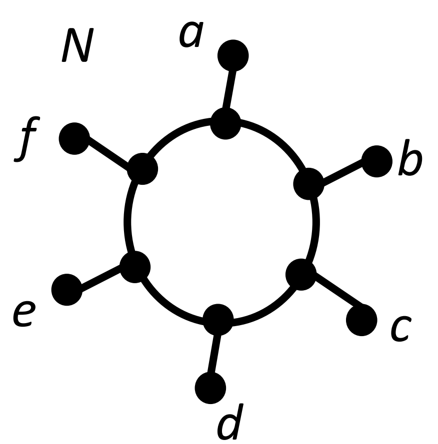


***N'=N* ?**

Reconstruction from triplets / quartets

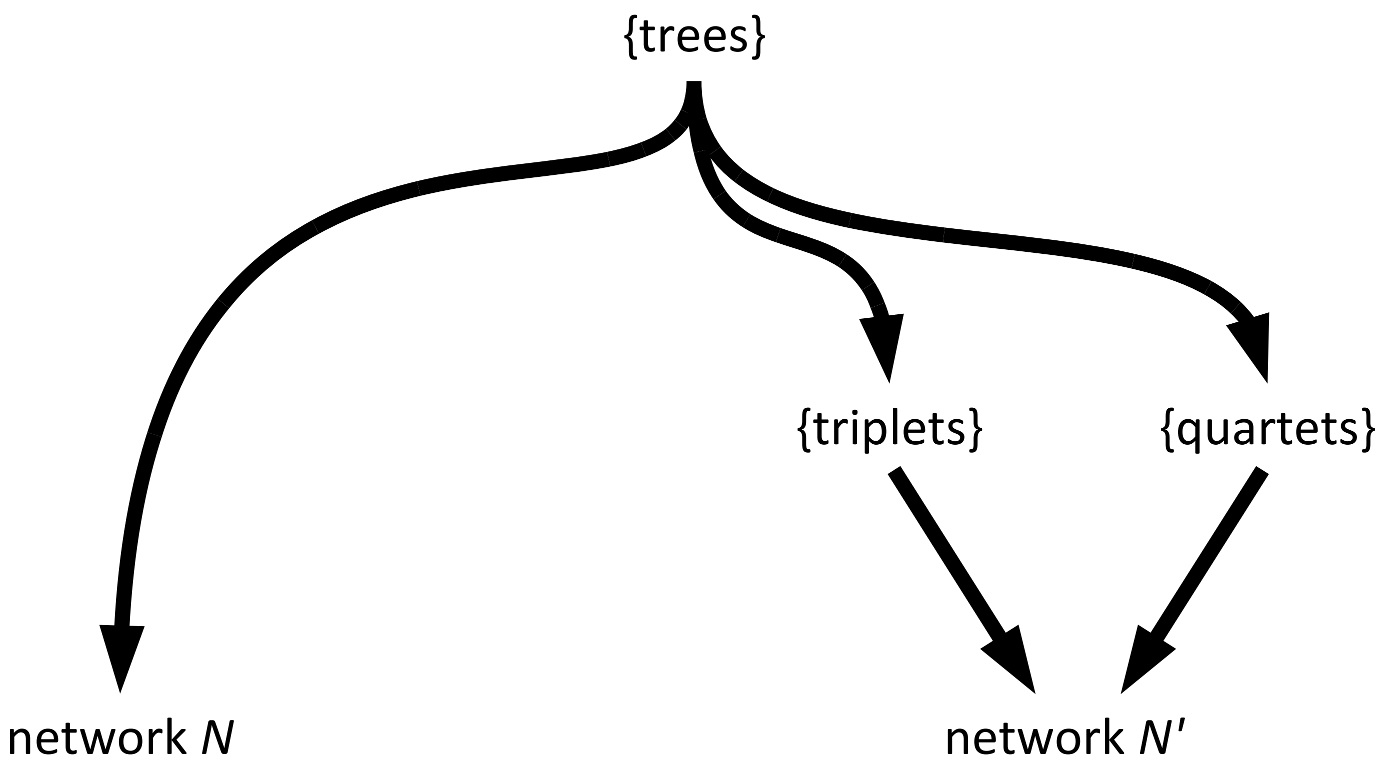
A network containing all quartets of a tree *T* **does not always** contain *T*.





N contains all quartets of *T* but not *T*

Reconstruction from triplets / quartets



$N'=N$?

Not always, but $\{N\} \subseteq \{N'\}$

Reconstruction from triplets / quartets

Finding **all triplets** of a rooted network: $O(n^3)$

Byrka, Gawrychowski, Huber & Kelk, *JDA*, 2010

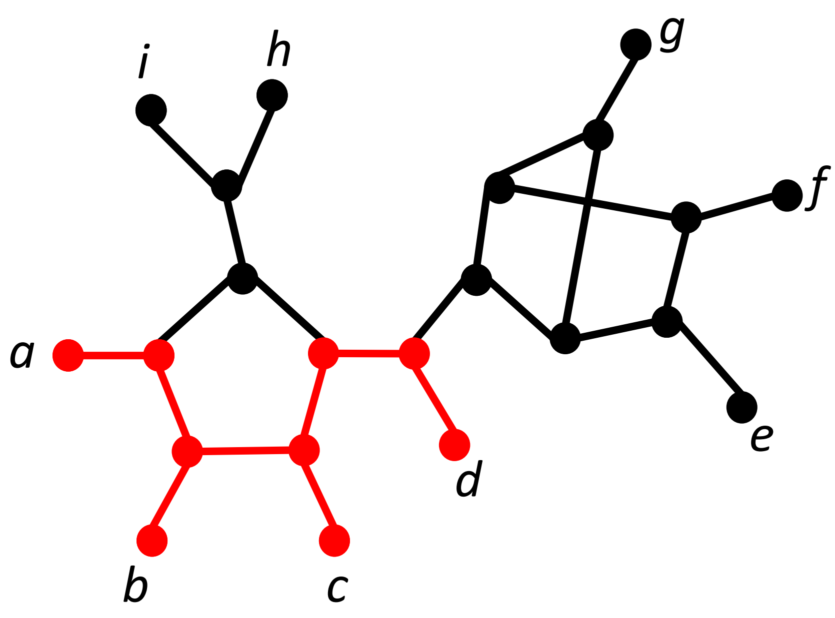
Reconstruction from triplets / quartets

Finding **all triplets** of a rooted network: $O(n^3)$

Byrka, Gawrychowski, Huber & Kelk, *JDA*, 2010

Finding **all quartets** of an unrooted network?

quartet *ab|cd*



Reconstruction from triplets / quartets

Finding **all triplets** of a rooted network: $O(n^3)$

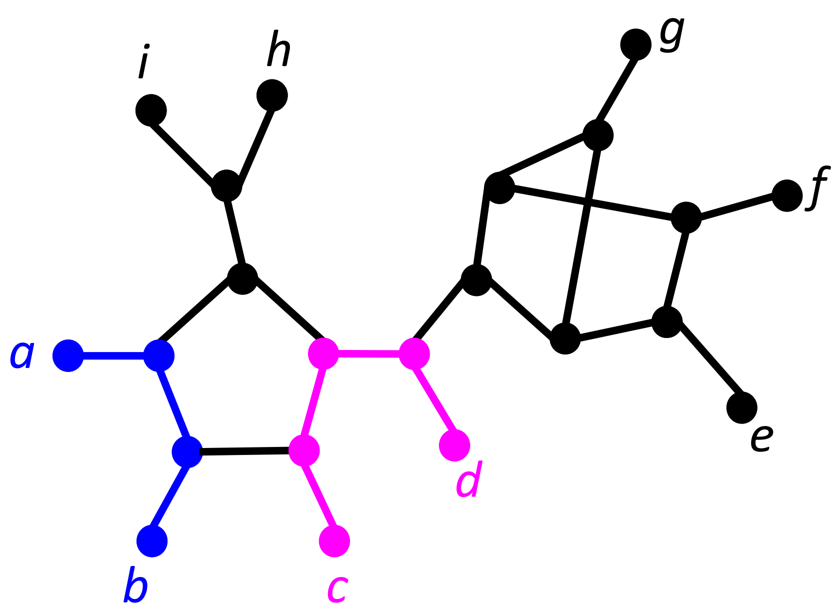
Byrka, Gawrychowski, Huber & Kelk, *JDA*, 2010

Finding **all quartets** of an unrooted network?

quartet *ab*|*cd*

⬍

2-disjoint paths
a-b,*c-d*



Reconstruction from triplets / quartets

Finding **all triplets** of a rooted network: $O(n^3)$

Byrka, Gawrychowski, Huber & Kelk, *JDA*, 2010

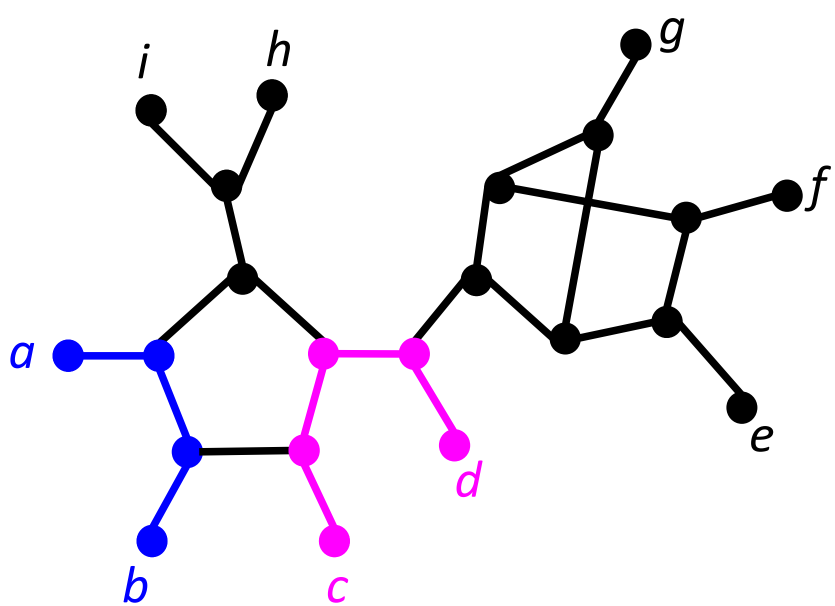
Finding **all quartets** of an unrooted network: $O(n^6)$
2-Disjoint Paths in a graph of degree ≤3: $O(n(1+\alpha(n,n)))$

Tholey, *SOFSEM'09*, 2009

quartet *ab*|*cd*

⬍

2-disjoint paths
a-b,*c-d*



Tree reconstruction

	unrooted from quartets	rooted from triplets
general	**NP-complete** Steel, *JOC*, 1992	**polynomial** Aho, Sagiv, Szymanski & Ullman, *SJOC*, 1981 Henzinger, King & Warnow, *ALG*, 1999 Jansson, Ng, Sadakane & Sung, *ALG*, 2005
dense *at least one quartet for each set of 4 leaves*	$O(n^4)$ Berry & Gascuel, *TCS*, 2000	$O(n^3)$ Aho et al., *SJOC*, 1981

Reconstruction of level-*k* networks

	unrooted from quartets		rooted from triplets	
	level 1	**level k>1**	**level 1**	**level k>1**
general	NP-complete Grünewald, Moulton & Spillner, *DAM*, 2009	?	NP-complete Jansson, Nguyen & Sung, *SJOC*, 2006	NP-complete Van Iersel, Kelk & Mnich, *JBCB*, 2009
dense *at least one quartet for each set of 4 leaves*	**?** **(decomposition in polynomial time)**	?	$O(n^3)$ Jansson, Nguyen & Sung, *SJOC*, 2006	$O(n^{5k+4})$ To & Habib, *CPM'09*
complete *all quartets of the network*	$O(n^4)$	? **(decomposition in polynomial time)**	$O(n^3)$ Jansson, Nguyen & Sung, *SJOC*, 2006	$O(n^{3k+3})$ Van Iersel & Kelk, *ALG*, 2010

Gambette, Berry & Paul, manuscript, 2011

Quartet set decomposition

$A|\overline{A}$ **SN-split** of the quartet set Q:
For all leaves $x,y \in A$, $z,t \in \overline{A}$,
The only quartet of Q on $\{x,y,z,t\}$ is $xy|zt$

Rooted context:
SN-set

Jansson & Sung, *TCS*, 2006
To & Habib, *CPM'09*

Q is a **dense** quartet set

⇓

SN-splits of Q are **compatible**
(can be represented by an unrooted tree)

Computing the SN-splits: $O(n^4)$

variant of the Q* algorithm
Berry & Gascuel, TCS, 2001

Rooted context:
Computing the SN-sets: $O(n^3)$

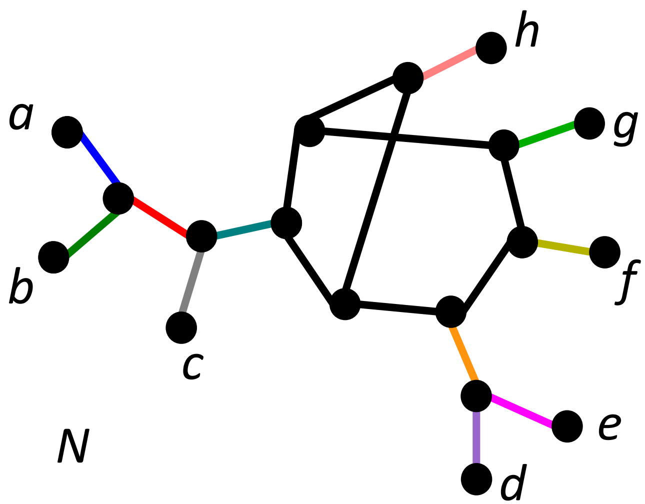
Jansson, Nguyen & Sung, *TCS*, 2006

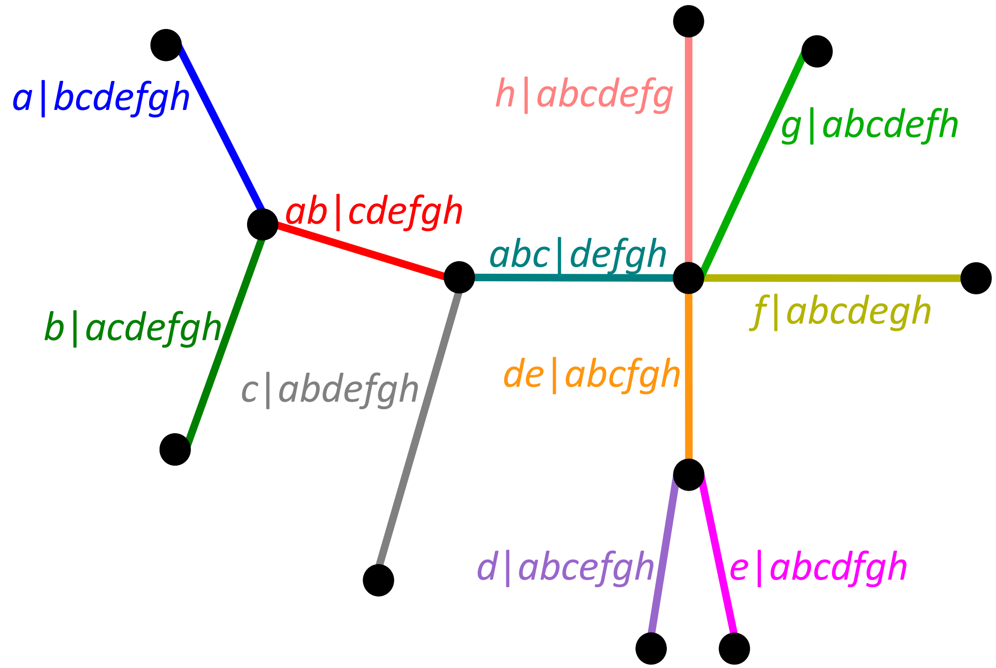
Gambette, Berry & Paul, manuscript, 2011

Quartet set decomposition

Q(*N*), the set of **all** quartets
contained in an unrooted level-*k* network *N*

SN-splits of *Q*(*N*)
are bijectively associated with **cut-edges** of *N.*



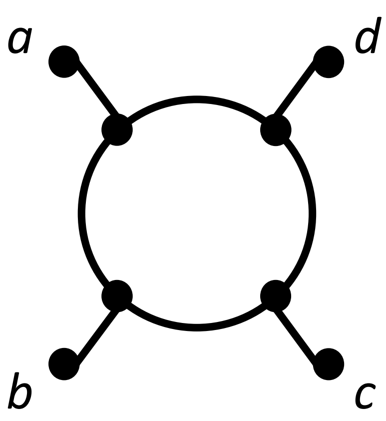


Gambette, Berry & Paul, manuscript, 2011

Blob reconstruction from quartets

We can separate the blobs of N from $Q(N)$.
How to **reconstruct** each blob of N?

Level-1 network reconstruction from the **complete quartet set:**
N, level-1 network
Finding an ordering of the leaves around the cycle: $O(n^2)$

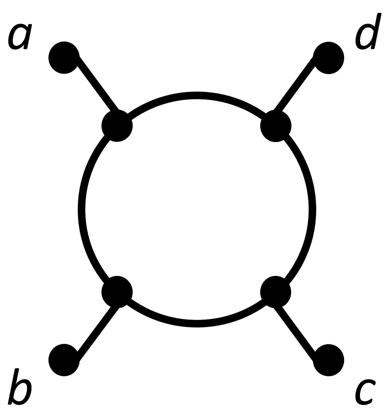


Quartet "circular puzzling"
Fix four leaves a,b,c,d:
$ab|cd$
$ad|bc$

Blob reconstruction from quartets

We can separate the blobs of N from $Q(N)$.
How to **reconstruct** each blob of N?

Level-1 network reconstruction from the **complete quartet set:**
N, level-1 network
Finding an ordering of the leaves around the cycle: $O(n^2)$

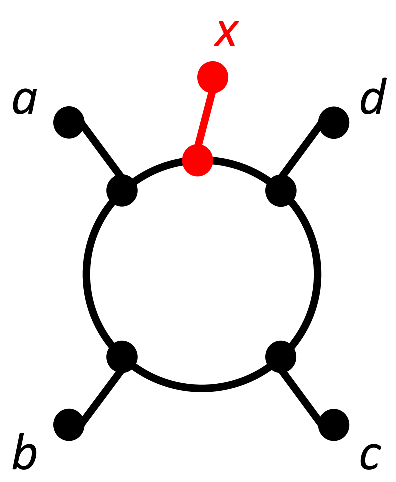


Quartet "circular puzzling"
Fix four leaves a,b,c,d ($ab|cd$, $ad|bc$)
For each other leaf:
- For each possible position on the cycle:
  - Test in $O(1)$ if it is the correct one.

Blob reconstruction from quartets

We can separate the blobs of N from $Q(N)$.
How to **reconstruct** each blob of N?

Level-1 network reconstruction from the **complete quartet set:**
N, level-1 network
Finding an ordering of the leaves around the cycle: $O(n^2)$



Quartet "circular puzzling"
Fix four leaves a,b,c,d ($ab|cd$, $ad|bc$)
For each other leaf:
- For each possible position on the cycle:
  - Test in $O(1)$ if it is the correct one.

$ba|xd + ax|db$?

Link with non-Betweenness

Betweenness:

Input:
- set X of elements
- set C of betweenness constraints:
  a between b and c

Output:
- ordering σ which respects the constraints of C

Link with non-Betweenness

Non-Betweenness:

Input:
- set X of elements
- set C of non-betweenness constraints:
  a not between b and c

Output:
- ordering σ which respects the constraints of C

Link with non-Betweenness

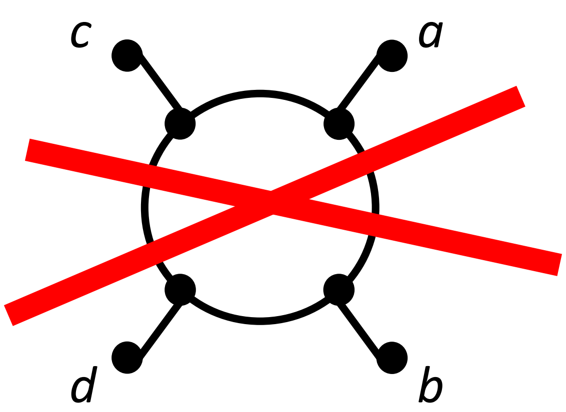
Circular Non-Betweenness:

Input:

- set *X* of elements
- set *C* of circular non-betweenness constraints:
  seen from *d*, *a* not between *b* and *c*

Output:

- circular ordering *σ* which respects the constraints of *C*



Link with non-Betweenness

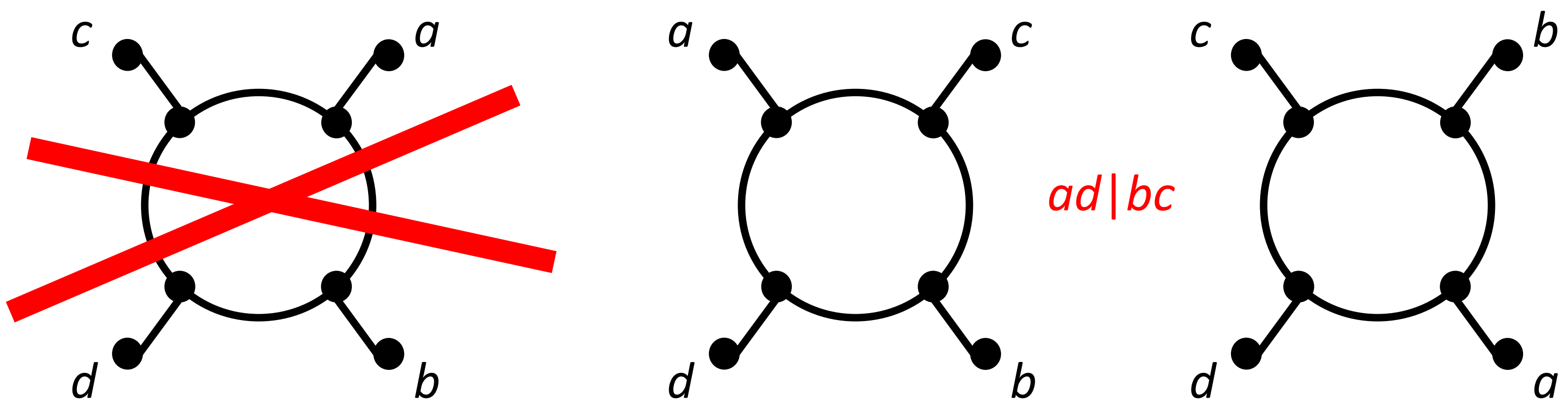
Circular Non-Betweenness:

Input:

- set *X* of elements
- set *C* of circular non-betweenness constraints:
  seen from *d*, *a* not between *b* and *c*

Output:

- circular ordering *σ* which respects the constraints of *C*



equivalent to unrooted level-1 blob reconstruction from quartets

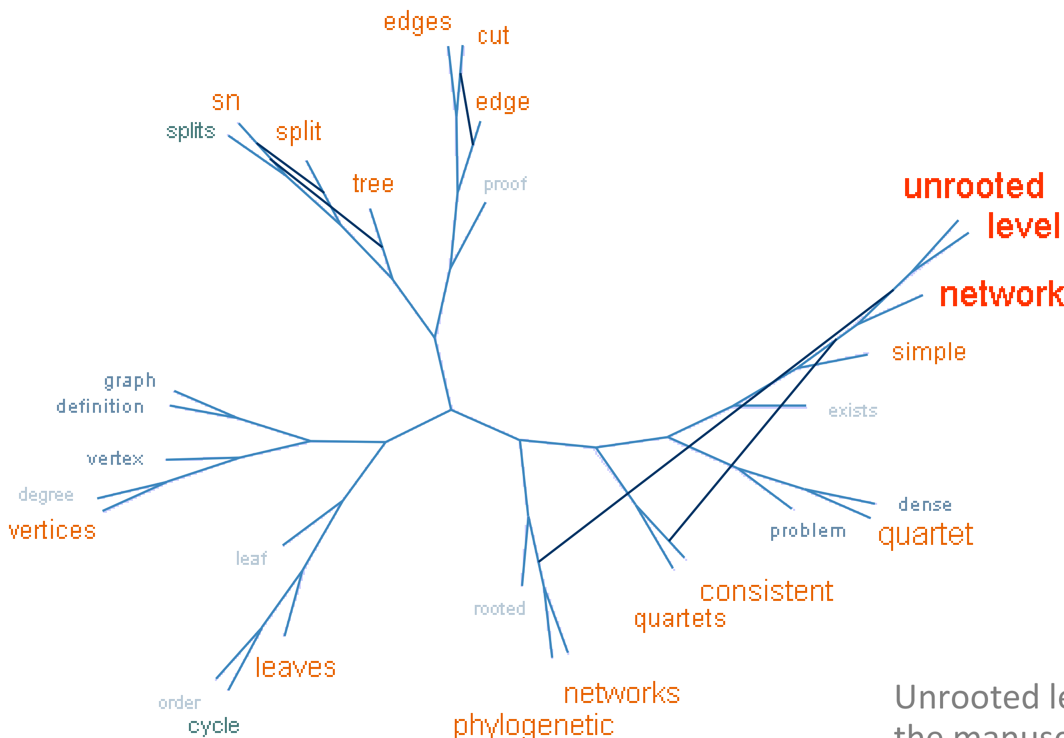
Link with non-Betweenness

	General case	Dense case
Betweenness	NP-complete Opatrny, *JOC*, 1979	polynomial
Non-Betweenness	NP-complete Guttman & Maucher, *IFIP-TCS'06*	?
Circular Non-Betweenness	NP-complete Grünewald, Moulton & Spillner, *DAM*, 2009	?

Thank you!

Coauthors of these results:

Vincent Berry & Christophe Paul (Montpellier), Mathilde Bouvel (Bordeaux)



Unrooted level-2 network built from the manuscript with

TreeCloud SplitsTree

T-Rex

http://www.treecloud.org