

AlgoB seminar, LIGM
26/05/2015 - Champs-sur-Marne

Finding a tree in a phylogenetic network

Philippe Gambette

Main results: joint work with Andreas Gunawan, Anthony Labarre, Stéphane Vialette and Louxin Zhang.



Outline

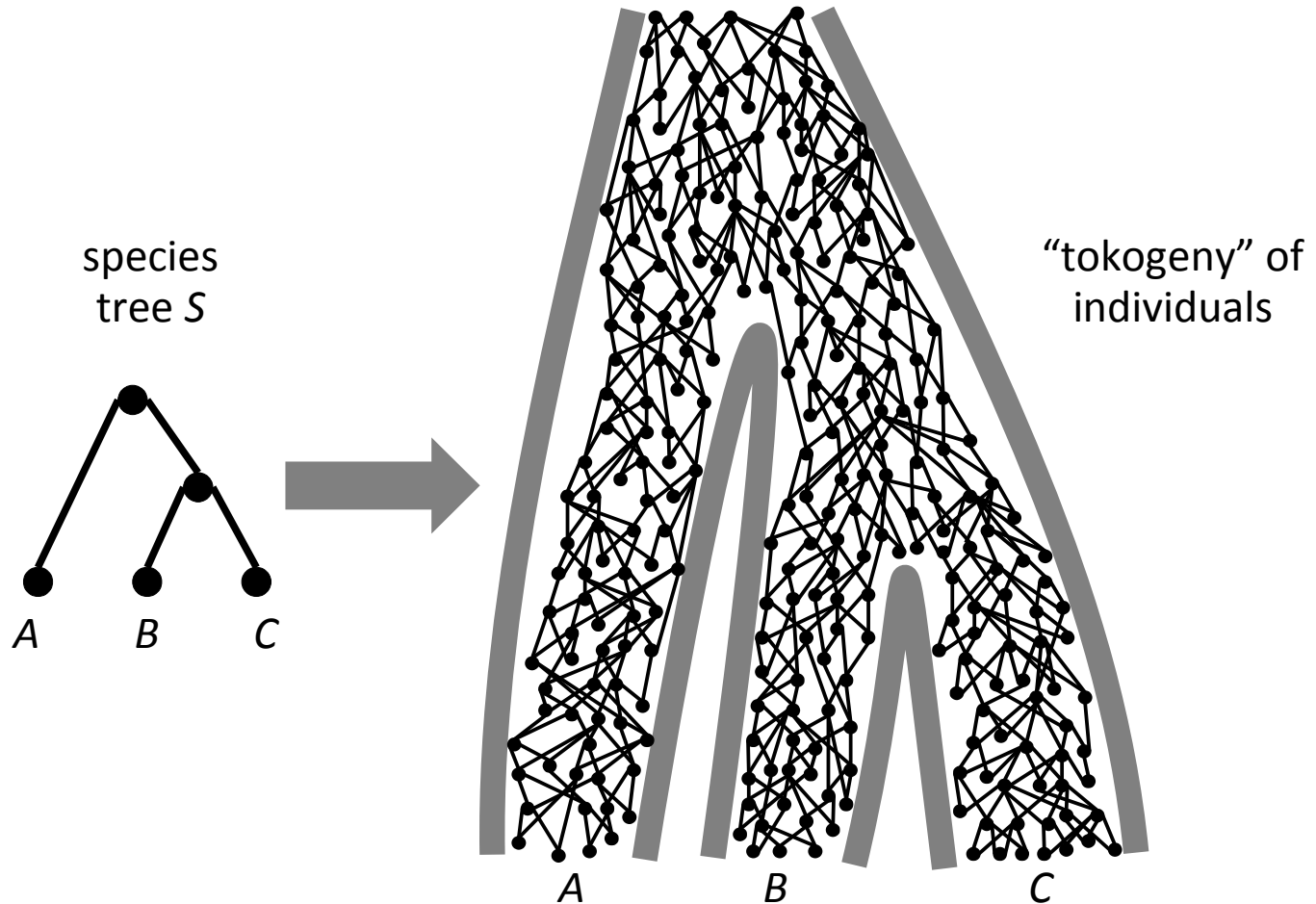
- Phylogenetic networks
- Who is who in phylogenetic networks?
- Classes of phylogenetic networks
- The Tree Containment Problem
- A quadratic-time algorithm on quasi-stable networks

Outline

- Phylogenetic networks
- Who is who in phylogenetic networks?
- Classes of phylogenetic networks
- The Tree Containment Problem
- A quadratic-time algorithm on quasi-stable networks

Phylogenetic trees

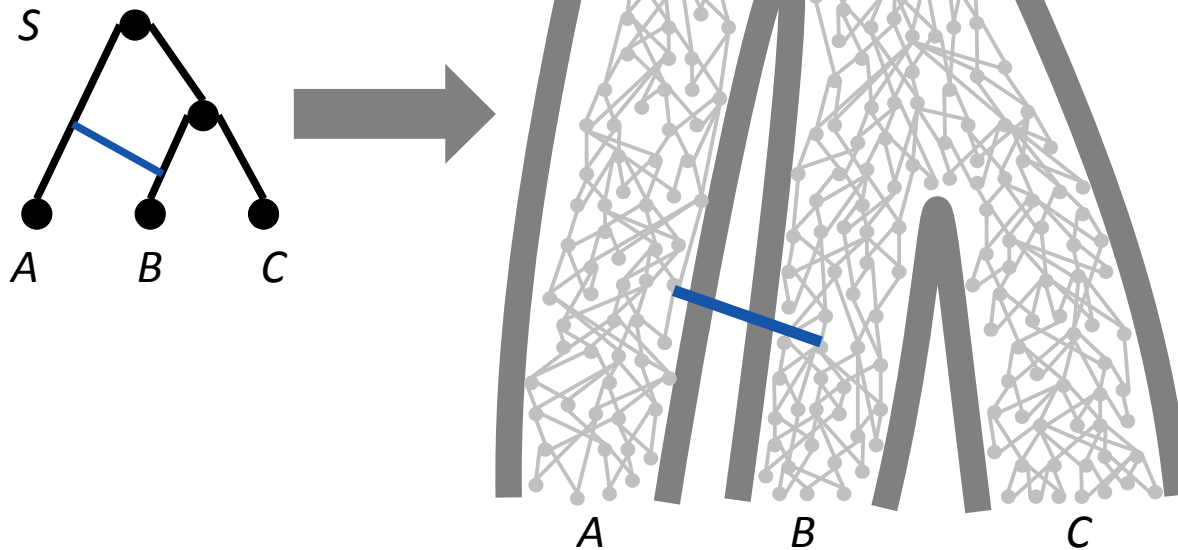
Phylogenetic tree of a set of species



Genetic material transfers

Transfers of genetic material between coexisting species:

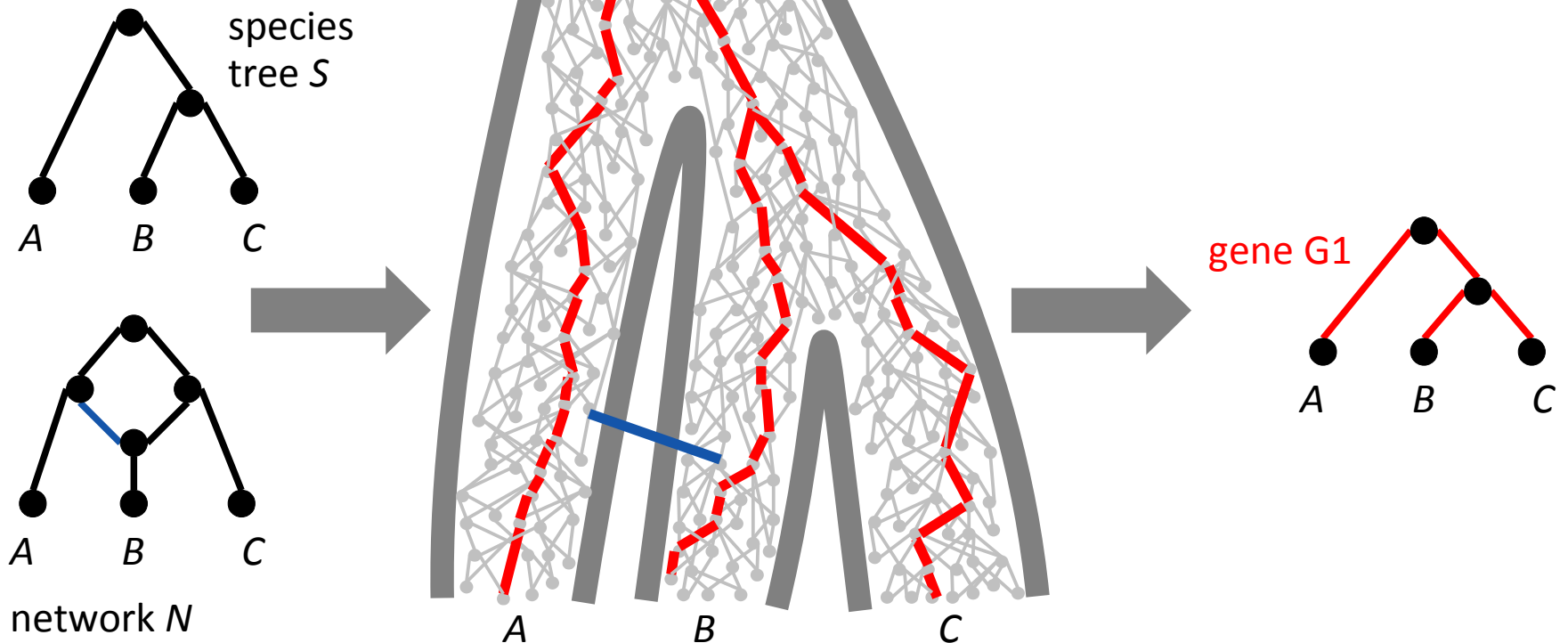
- lateral gene transfer
- hybridization
- recombination



Genetic material transfers

Transfers of genetic material between coexisting species:

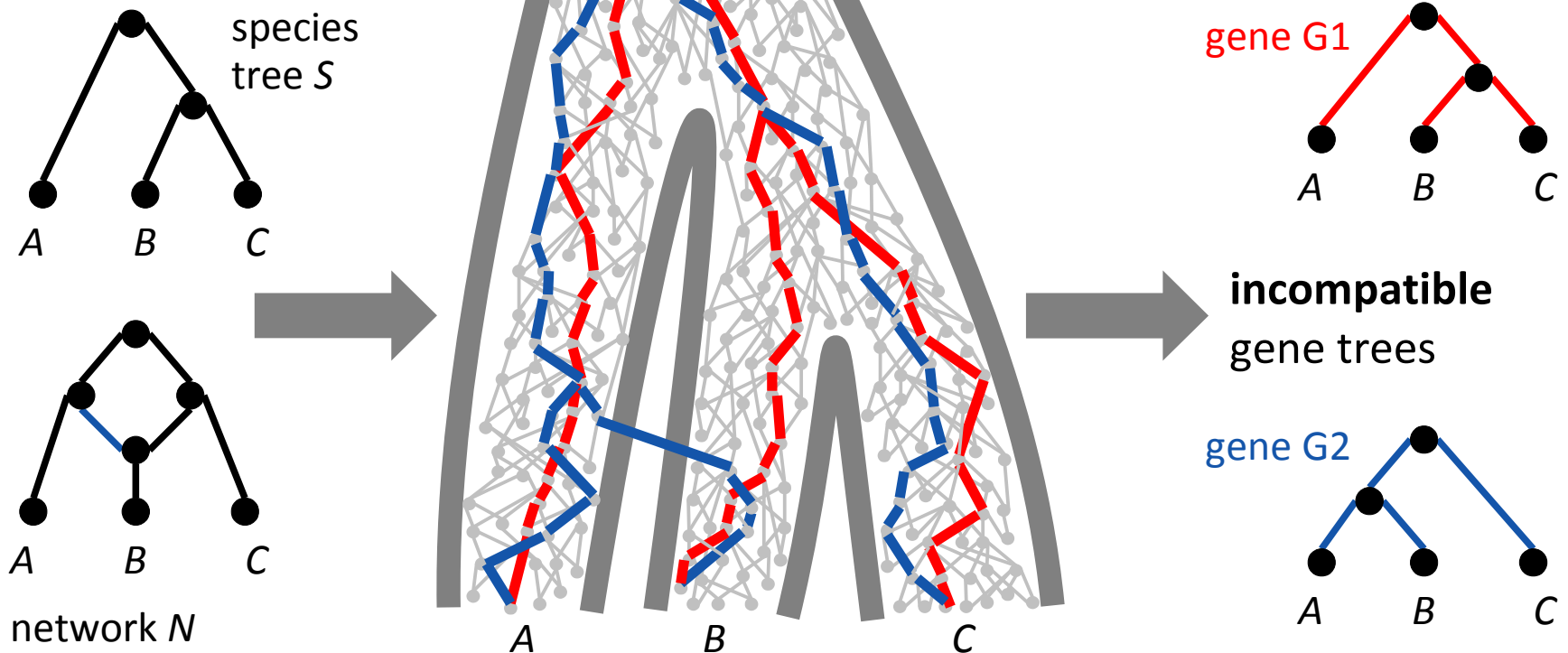
- lateral gene transfer
- hybridization
- recombination



Genetic material transfers

Transfers of genetic material between coexisting species:

- lateral gene transfer
- hybridization
- recombination

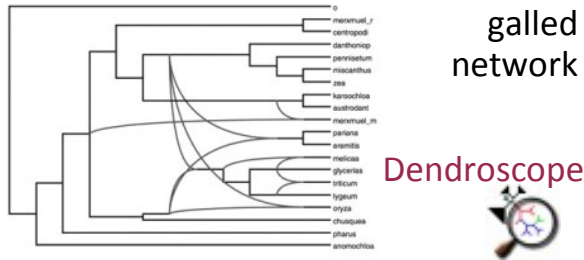


Phylogenetic networks

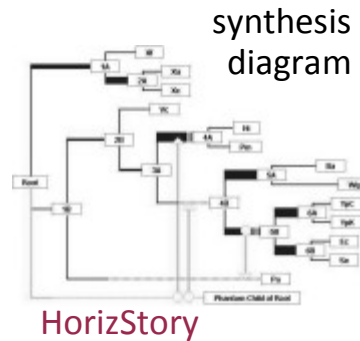
Phylogenetic network: network representing evolution data

- **explicit** phylogenetic networks

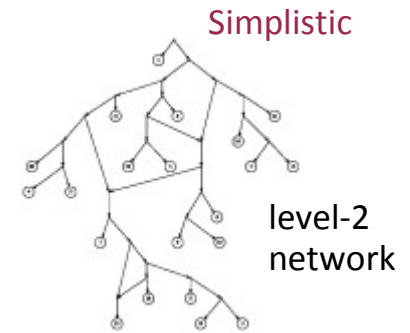
model evolution



Dendroscope



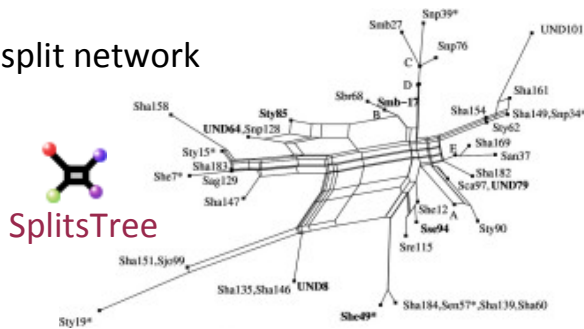
HorizStory



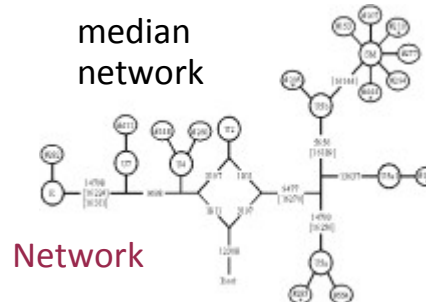
- **abstract** phylogenetic networks

classify, visualize data

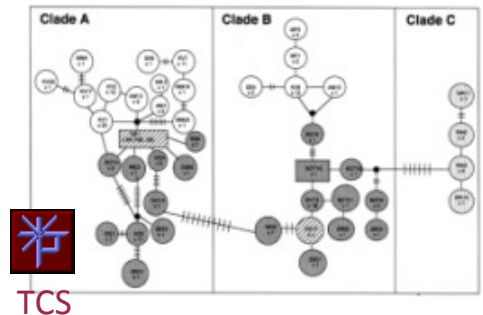
split network



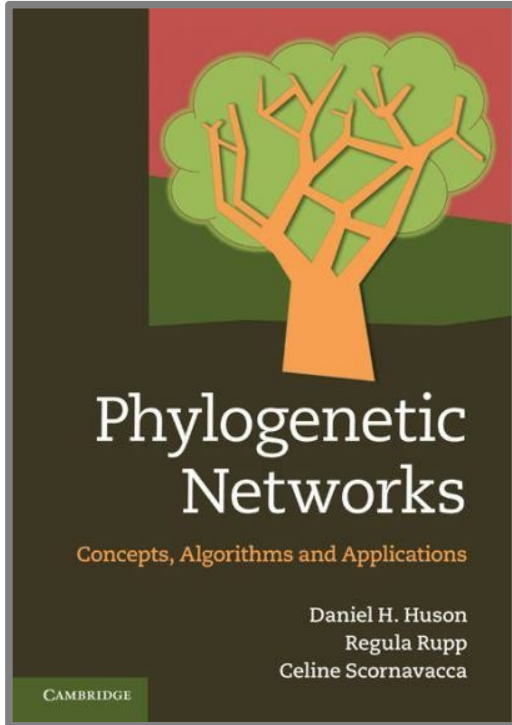
median network



minimum spanning network

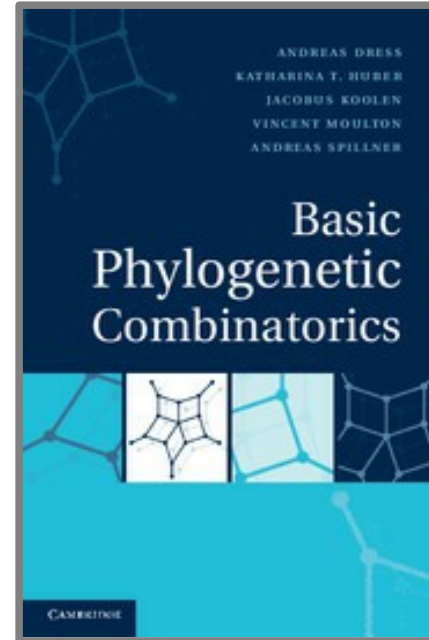


Books about phylogenetic networks

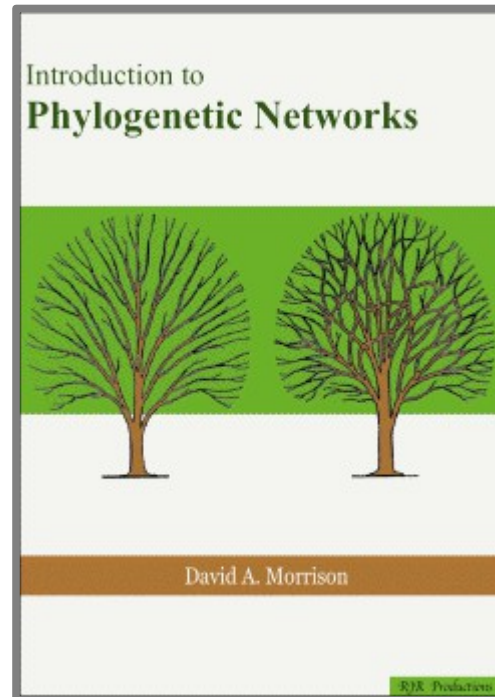


Huson, Rupp,
Scornavacca, 2011

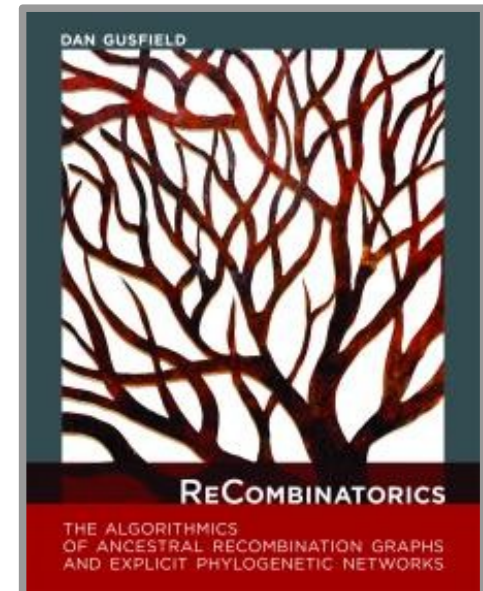
Dress, Huber,
Koolen, Moulton,
Spillner, 2012



Gusfield, 2014



Morrison, 2011



Workshops about phylogenetic networks



The Future of Phylogenetic Networks, 15-19 October **2012**, Lorentz Center, Leiden, The Netherlands

Utilizing Genealogical Phylogenetic Networks in Evolutionary Biology: Touching the Data, 7-11 July **2014**, Lorentz Center, Leiden, The Netherlands

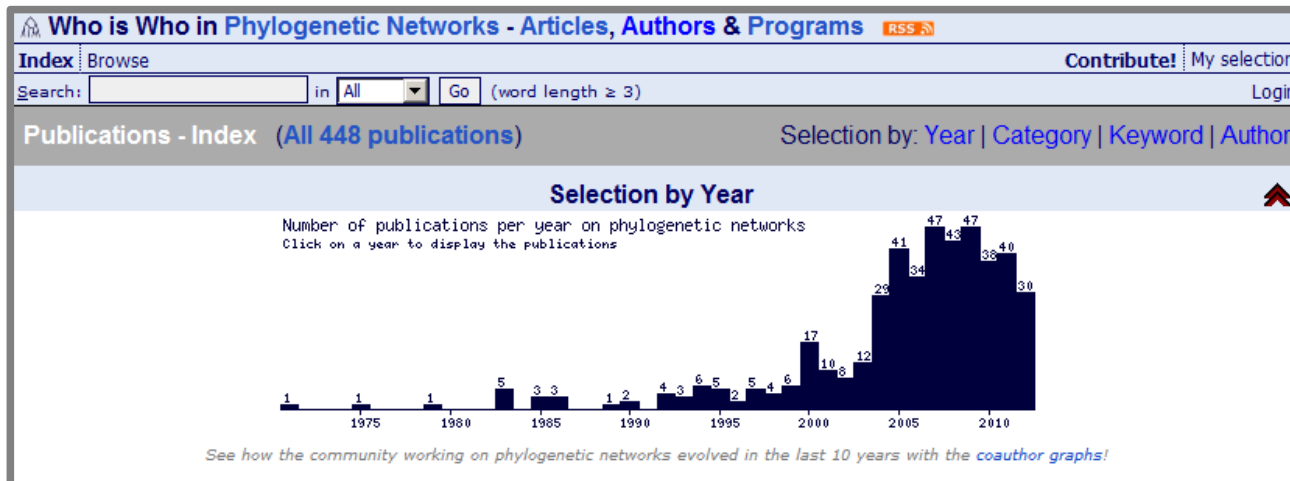


The Phylogenetic Network Workshop, 27 - 31 Jul **2015**, Institute for Mathematical Science (National University of Singapore)

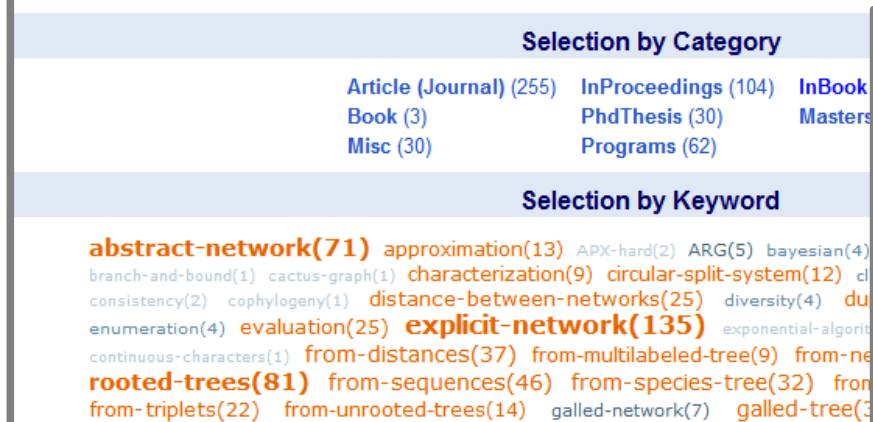
Outline

- Phylogenetic networks
- Who is who in phylogenetic networks?
- Classes of phylogenetic networks
- The Tree Containment Problem
- A quadratic-time algorithm on quasi-stable networks

Who is who in Phylogenetic Networks?



Who is Who in
Phylogenetic
Networks, **Articles,**
Authors &
Programs



Who is Who in Phylogenetic Networks - Articles, Authors & Programs

Index Browse Search: [] in All Go (word length ≥ 3) Contribute! My selection Login

Programs to compute, evaluate, compare, visualize... **phylogenetic networks**
This page is automatically built from all publications tagged by Program* in the [database](#).

Program Arlequin

The goal of *Arlequin* is to provide the average user in population genetics with quite a large set of basic methods and statistical tests, in order to extract information on genetic and demographic features of a collection of population samples. In particular, Arlequin implements a Minimum Spanning Network algorithm to embed the set of all minimum spanning trees computed from a distance matrix of haplotypes (<http://cmpg.unibe.ch/software/arlequin3/>).

5 publications in the database mention Program Arlequin

Program Beagle

Beagle is a small collection of related programs for analysing the minimum number of recombinations required for a SNP data set under the infinite sites model. Available at <http://www.stats.ox.ac.uk/~lyngsoe/beagle/>.

3 publications in the database mention Program Beagle

Program Bio-PhyloNetwork

Bio-PhyloNetwork is a Perl package that relies on the BioPerl bundle and implements many algorithms on phylogenetic networks (<http://dmi.uib.es/~gcardona/BioInfo/Bio-PhyloNetwork.tgz>). It is used in a Java Applet which can compare and draw two phylogenetic networks entered in eNewick format with the same set of leaves (<http://dmi.uib.es/~gcardona/BioInfo/alignment.php>)

4 publications in the database mention Program Bio-PhyloNetwork

Based on BibAdmin by Sergiu Chelcea + tag clouds, date histograms, journal lists, keyword definitions, co-author graphs

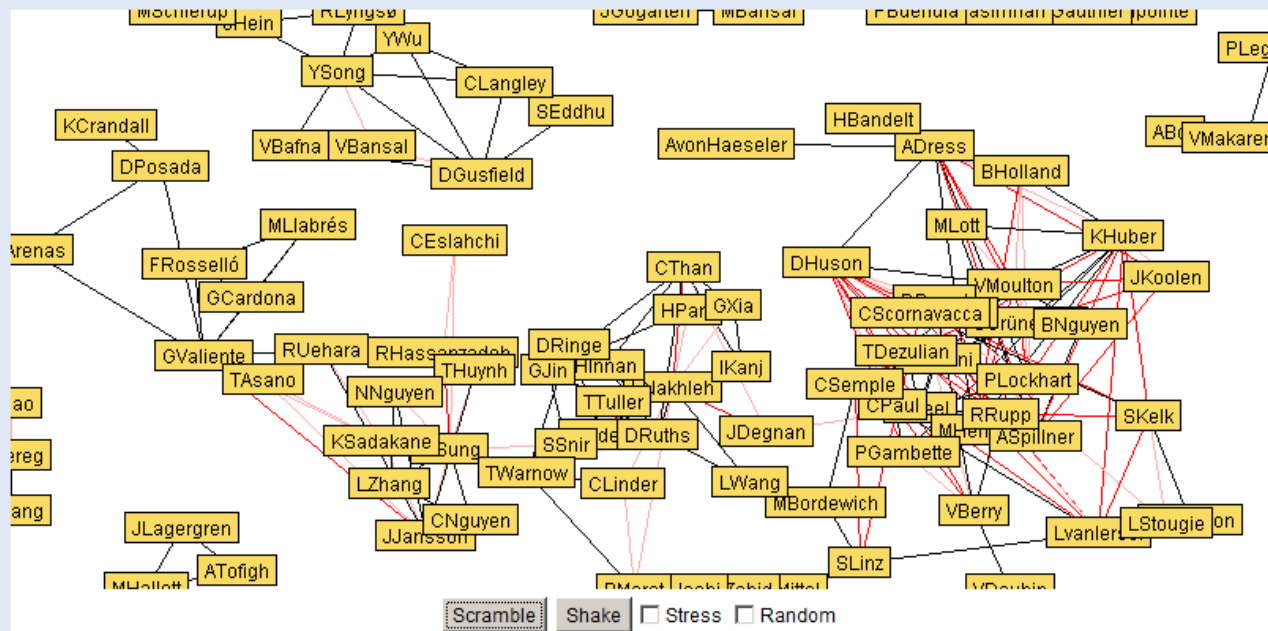
Analysis of the co-author and keyword graphs: internship of **Tushar Agarwal**

Who is who in Phylogenetic Networks?

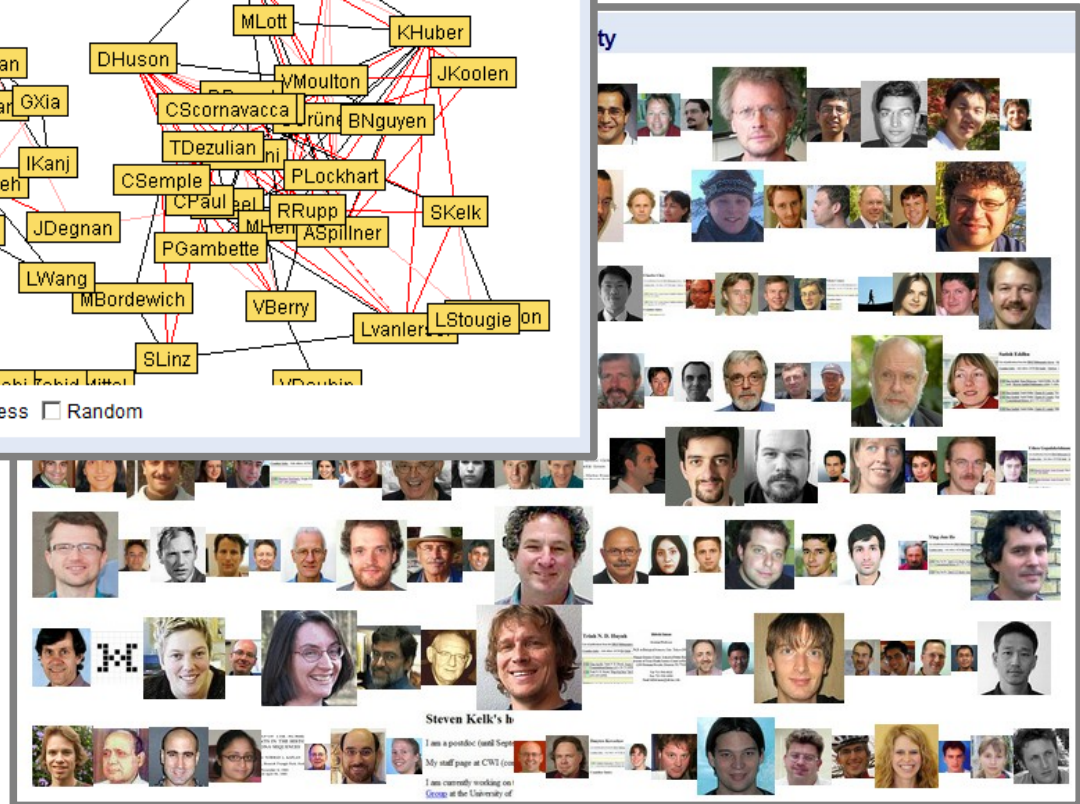
Coauthor Graphs

The graphs below link authors of at least two publications who published on phylogenetic networks at least twice together. They show how the community of mathematicians and computer scientists working on phylogenetic networks evolved and structured itself over the last 10 years.

Coauthor graph today

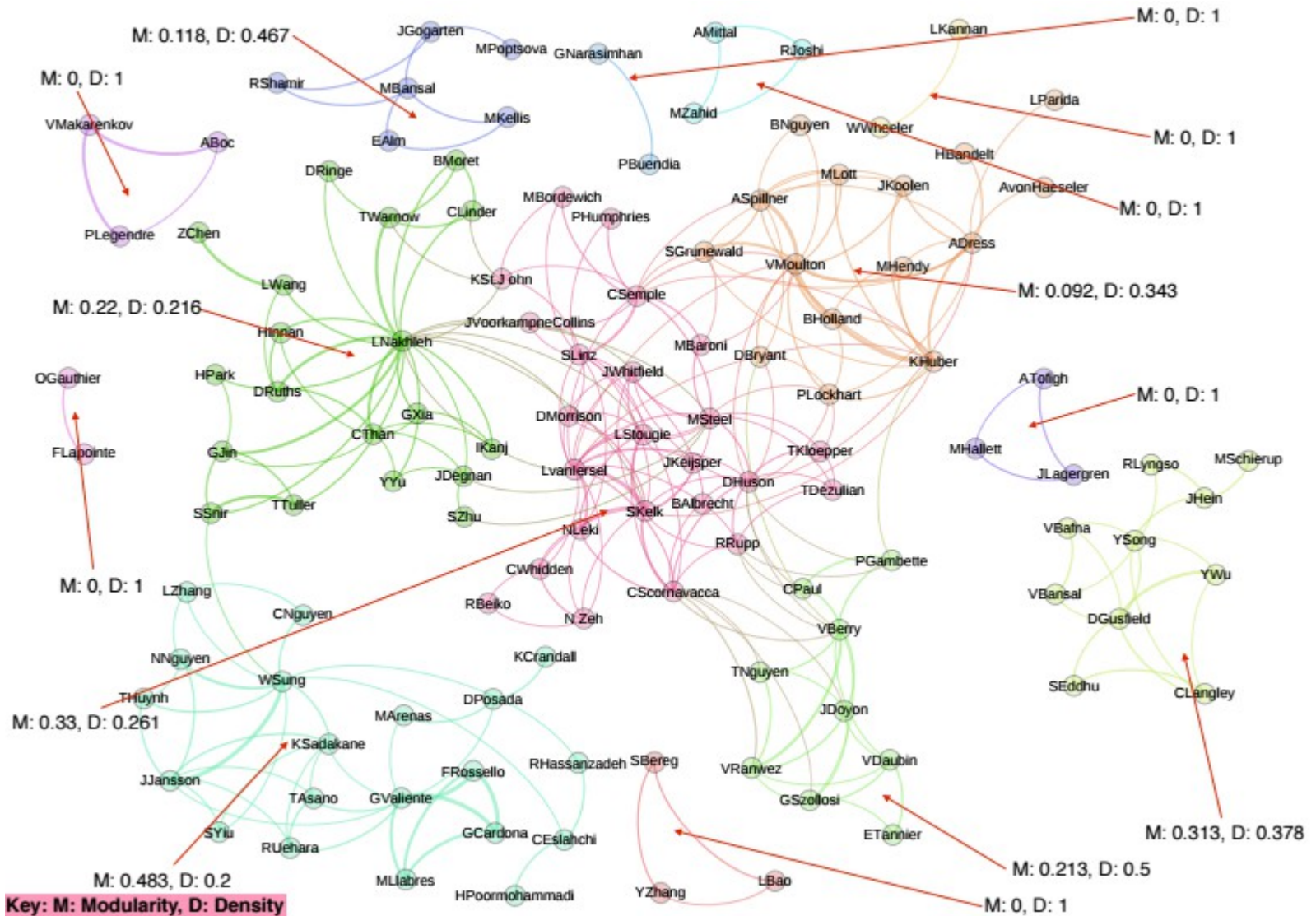


Who is Who in
Phylogenetic
Networks, Articles,
**Authors &
Programs**



Analysis of the co-author and
keyword graphs: internship of
Tushar Agarwal

Who is who in Phylogenetic Networks?




Gephi visualization of the coauthor network of main authors by Tushar


Who is who in Phylogenetic Networks?

Who is Who in Phylogenetic Networks - Articles, Authors & Programs RSS


Index: **Browse** Contribute! My selection


Search: in **All** Go (word length \geq 3) Login


Publications of **Daniel H. Huson**   Order by: Type | Year

Associated keywords 


abstract-network circular-split-system consensus **explicit-network** FPT from-clusters from-distances from-network from-rooted-trees from-sequences **from-splits** from-trees from-unrooted-trees galled-network galled-tree heuristic hybridization level-k-phylogenetic-network minimum-number NP-complete **phylogenetic-network** **phylogeny** polynomial Program-Beagle **Program-Dendroscope** Program-HybridInterleave Program-HybridNumber Program-Spectronet **Program-SplitsTree** Program-SPNet recombination **reconstruction software** split split-decomposition **split-network** supernetwork survey tanglegram **visualization**


<< 2012 >> 

1  

[Benjamin Albrecht](#), [Celine Scornavacca](#), [Alberto Cenci](#) and [Daniel H. Huson](#). Fast computation of minimum hybridization networks. In *Bioinformatics*, Vol. 28(2):191-197, 2012. [Comment]  

Keywords: explicit network, from rooted trees, minimum number, phylogenetic network, phylogeny, Program Dendroscope, reconstruction. **Note:** <http://dx.doi.org/10.1093/bioinformatics/btr618>.


<< 2011 >> 

2 

Basé sur BibAdmin
par Sergiu Chelcea
+ nuages de mots, histogramme
des dates, liste des journaux,
graphes de co-auteurs,
définition des mots-clés.

Who is Who in
Phylogenetic
Networks, **Articles,**
Authors &
Programs

Community



Steven Kelk's h
I am a postdoc (until Sep...
My staff page at CWI (com...
I am currently working on...
Group at the University of

Who is who in Phylogenetic Networks?

abstract-network(88) agreement-forest(29) approximation(18) APX-hard(4) ARG(6) bayesian(4) block-realization(1) bootstrap(4) bound(4) branch-and-bound(1) cactus-graph(1) characterization(11) circular-split-system(13) clustering(3) coalescent(10) consensus(8) consistency(2) cophylogeny(1) counting(3) database(1) distance-between-networks(30) diversity(5) duplication(33) dynamic-programming(10) enumeration(4) **evaluation(25)** **explicit-network(191)** exponential-algorithm(6) FPT(31) from-clusters(16) from-continuous-characters(1) from-distances(45) from-gene-order(1) from-multilabeled-tree(9) from-network(39) from-NGS-data(1) from-quartets(16) **from-rooted-trees(122)** from-sequences(51) from-species-tree(46) from-splits(13) from-trees(7) from-trinets(2) from-triplets(24) from-unrooted-trees(16) galled-network(7) galled-tree(38) generation(12) haplotype-network(2) haplotyping(1) heuristic(32) HMM(2) hybridization(52) inapproximability(5) integer-linear-programming(3) isomorphism(3) k-reticulated(1) kernelization(2) labeling(4) lateral-gene-transfer(56) level-k-phylogenetic-network(30) likelihood(16) lineage-sorting(10) loss(16) MASN(4) median-network(16) MedianJoining(2) Minimal-lateral-network(1) minimum-contradiction(2) **minimum-number(33)** minimum-spanning-network(2) model-selection(2) mu-distance(2) **NeighborNet(14)** nested-network(2) netting(3) normal-network(9) NP-complete(33) optimal-realization(3) parsimony(39) pedigree(1) perfect(5) **phylogenetic-network(402)** **phylogeny(387)** **polynomial(70)** population-genetics(5) Program-AdmixTools(1) Program-ALE(1) Program-Angst(2) Program-Arlequin(5) Program-Beagle(3) Program-BIMLR(1) Program-Bio-PhyloNetwork(4) Program-Clustistic(2) Program-CMPT(1) Program-CombineTrees(2) Program-ConsensusNetwork(1) Program-constNJ(1) Program-CycleKiller(3) **Program-Dendroscope(14)** Program-EEEP(3) Program-FastHN(1) Program-FlatNJ(1) Program-Fylogenetica(2) Program-GalledTree(1) Program-GraphDTL(2) Program-HapBound(1) Program-HGT_simul(1) Program-HiDe(1) Program-HorizStory(2) Program-Hybrid-Lambda(1) Program-HybridInterleave(5) Program-HybridNET(2) Program-HybridNumber(3) Program-Hybroscale(3) Program-JML(1) Program-LatTrans(5) Program-LEV1ATHAN(1) Program-Level1Generator(1) Program-Level2(2) Program-lingpy(1) Program-LNetwork(2) Program-MaafB(2) Program-Marlon(3) Program-MC-Net(1) Program-McKITsch(1) Program-Mowgli(5) Program-MowgliNNI(2) Program-MPNet(1) Program-MY-CLOSURE(1) **Program-Nepal(7)** Program-NetGen(3) Program-NetTest(1) Program-NetView(1) Program-Network(5) Program-Notung(1) **Program-PADRE(7)** Program-Phangorn(2) Program-PhippsNetwork(2) **Program-PhyloNet(11)** Program-PhyloNet-HMM(1) Program-PIRN(3) Program-Prunier(2) Program-Pyramids(3) Program-QNet(4) Program-Quartet(1) Program-Quartet-Decomposition(1) Program-QuartetMethods(1) Program-QuartetNet(2) Program-QuasiDec(1) Program-QuickCass(1) Program-RANGER-DTL(4) Program-RecMin(1) Program-Recodon(3) Program-RecPars(1) Program-Reticlad(2) Program-SAGE(2) Program-SAQ-Net(1) Program-Serial-NetEvolve(1) Program-SHRUB(3) Program-Simplistic(3) Program-Sliding-MinPD(1) Program-SNSA(2) Program-Spectronet(4) **Program-SplitsTree(36)** Program-SPNet(5) Program-SPRDist(1) Program-SuperQ(1) **Program-T-REX(13)** Program-TCS(8) Program-TERA(2) Program-TerminusEst(2) Program-TreeFix-DTL(1) Program-TreeMix(1) Program-Treevolve(2) Program-TripNet(2) Program-ultra-Net(1) Program-Ultranet(1) Program-WeakHierarchies(2) Program-Xscape(1) pyramid(8) quasi-median-network(3) realization(4) **recombination(29)** recombination-detection(4) **reconstruction(255)** regular-network(7) reticulogram(10) serial-evolutionary-networks(1) simulated-annealing(4) simulation(5) site-consistency(1) **software(66)** split(27) split-decomposition(13) split-network(54) SPR-distance(13) spread(2) **statistical-model(32)** statistical-parsimony(3) supernet(4) **survey(31)** tanglegram(1) time-consistent-network(12) tree-child-network(15) tree-sibling-network(10) tripartition-distance(9) triplet-distance(3) unicyclic-network(3) **visualization(30)** weak-hierarchy(8) weakly-compatible(3)

Who is who in Phylogenetic Networks?

abstract-network(88) agreement-forest(29) approximation(18) APX-hard(4) ARG(6) bayesian(4) block-realization(1) bootstrap(4) bound(4) branch-and-bound(1) cactus-graph(1) characterization(11) circular-split-system(13) clustering(3) coalescent(10) consensus(8) consistency(2) cophylogeny(1) counting(3) database(1) distance-between-networks(30) diversity(5) duplication(33) dynamic-programming(10) enumeration(4) **evaluation(25)** **explicit-network(191)** exponential-algorithm(6) **FPT(31)** from-clusters(16) from-continuous-characters(1) from-distances(45) from-gene-order(1) from-multilabeled-tree(9) from-network(39) from-NGS-data(1) from-quartets(16) **from-rooted-trees(122)** from-sequences(51) from-species-tree(46) from-splits(13) from-trees(7) from-trinets(2) from-triplets(24) from-unrooted-trees(16) galled-network(7) galled-tree(38) generation(12) haplotype-network(2) haplotyping(1) heuristic(32) HMM(2) hybridization(52) inapproximability(5) integer-linear-programming(3) isomorphism(3) k-reticulated(1) kernelization(2) labeling(4) lateral-gene-transfer(56) level-k-phylogenetic-network(30) likelihood(16) lineage-sorting(10) loss(16) MASN(4) median-network(16) MedianJoining(2) Minimal-lateral-network(1) minimum-contradiction(2) **minimum-number(33)** minimum-spanning-network(2) model-selection(2) mu-distance(2) NeighborNet(14) nested-network(2) netting(3) normal-network(9) NP-complete(33) optimal-realization(3) parsimony(39) pedigree(1) perfect(5) **phylogenetic-network(402)** **phylogeny(387)** **polynomial(70)** population-genetics(5) Program-AdmixTools(1) Program-ALE(1) Program-Angst(2) Program-Arlequin(5) Program-Beagle(3) Program-BIMLR(1) Program-Bio-PhyloNetwork(4) Program-Clustistic(2) Program-CMPT(1) Program-CombineTrees(2) Program-ConsensusNetwork(1) Program-constNJ(1) Program-CycleKiller(3) **Program-Dendroscope(14)** Program-EEEP(3) Program-FastHN(1) Program-FlatNJ(1) Program-Fylogenetica(2) Program-GalledTree(1) Program-GraphDTL(2) Program-HapBound(1) Program-HGT_simul(1) Program-HiDe(1) Program-HorizStory(2) Program-Hybrid-Lambda(1) Program-HybridInterleave(5) Program-HybridNET(2) Program-HybridNumber(3) Program-Hybroscale(3) Program-JML(1) Program-LatTrans(5) Program-LEV1ATHAN(1) Program-Level1Generator(1) Program-Level2(2) Program-lingpy(1) Program-LNetwork(2) Program-MaafB(2) Program-Marlon(3) Program-MC-Net(1) Program-McKITsch(1) Program-Mowgli(5) Program-MowgliNNI(2) Program-MPNet(1) Program-MY-CLOSURE(1) Program-Nepal(7) Program-NetGen(3) Program-NetTest(1) Program-NetView(1) Program-Network(5) Program-Notung(1) Program-PADRE(7) Program-Phangorn(2) Program-PhippsNetwork(2) **Program-PhyloNet(11)** Program-PhyloNet-HMM(1) Program-PIRN(3) Program-Prunier(2) Program-Pyramids(3) Program-QNet(4) Program-Quartet(1) Program-Quartet-Decomposition(1) Program-QuartetMethods(1) Program-QuartetNet(2) Program-QuasiDec(1) Program-QuickCass(1) Program-RANGER-DTL(4) Program-RecMin(1) Program-Recodon(3) Program-RecPars(1) Program-Reticlad(2) Program-SAGE(2) Program-SAQ-Net(1) Program-Serial-NetEvolve(1) Program-SHRUB(3) Program-Simplistic(3) Program-Sliding-MinPD(1) Program-SNSA(2) Program-Spectronet(4) **Program-SplitsTree(36)** Program-SPNet(5) Program-SPRDist(1) Program-SuperQ(1) **Program-T-REX(13)** Program-TCS(8) Program-TERA(2) Program-TerminusEst(2) Program-TreeFix-DTL(1) Program-TreeMix(1) Program-Treevolve(2) Program-TripNet(2) Program-ultra-Net(1) Program-Ultranet(1) Program-WeakHierarchies(2) Program-Xscape(1) pyramid(8) quasi-median-network(3) realization(4) recombination(29) recombination-detection(4) **reconstruction(255)** regular-network(7) reticulogram(10) serial-evolutionary-networks(1) simulated-annealing(4) simulation(5) site-consistency(1) **software(66)** split(27) split-decomposition(13) split-network(54) SPR-distance(13) spread(2) statistical-model(32) statistical-parsimony(3) supernet(4) survey(31) tanglegram(1) time-consistent-network(12) tree-child-network(15) tree-sibling-network(10) tripartition-distance(9) triplet-distance(3) unicyclic-network(3) visualization(30) weak-hierarchy(8) weakly-compatible(3)

input

software

Who is who in Phylogenetic Networks?

abstract-network(88) agreement-forest(29) approximation(18) APX-hard(4) ARG(6) bayesian(4) block-realization(1) bootstrap(4) bound(4) branch-and-bound(1) cactus-graph(1) characterization(11) circular-split-system(13) clustering(3) coalescent(10) consensus(8) consistency(2) cophylogeny(1) counting(3) database(1) distance-between-networks(30) diversity(5) duplication(33) dynamic-programming(10) enumeration(4) evaluation(25) **explicit-network(191)** exponential-algorithm(6) **FPT(31)** from-clusters(16) from-continuous-characters(1) from-distances(45) from-gene-order(1) from-multilabeled-tree(9) from-network(39) from-NGS-data(1) from-quartets(16) **from-rooted-trees(122)** from-sequences(51) from-species-tree(46) from-splits(13) from-trees(7) from-trinets(2) from-triplets(24) from-unrooted-trees(16) galled-network(7) **galled-tree(38)** generation(12) haplotype-network(2) haplotyping(1) heuristic(32) HMM(2) hybridization(52) inapproximability(5) integer-linear-programming(3) isomorphism(3) k-reticulated(1) kernelization(2) labeling(4) lateral-gene-transfer(56) **level-k-phylogenetic-network(30)** likelihood(16) lineage-sorting(10) loss(16) MASN(4) **median-network(16)** MedianJoining(2) Minimal-lateral-network(1) minimum-contradiction(2) **minimum-number(33)** minimum-spanning-network(2) model-selection(2) mu-distance(2) NeighborNet(14) **nested-network(2)** netting(3) **normal-network(9)** **NP-complete(33)** optimal-realization(3) parsimony(39) pedigree(1) perfect(5) **phylogenetic-network(402)** **phylogeny(387)** **polynomial(70)** population-genetics(5) Program-AdmixTools(1) Program-ALE(1) Program-Angst(2) Program-Arlequin(5) Program-Beagle(3) Program-BIMLR(1) Program-Bio-PhyloNetwork(4) Program-Clustistic(2) Program-CMPT(1) Program-CombineTrees(2) Program-ConsensusNetwork(1) Program-constNJ(1) Program-CycleKiller(3) **Program-Dendroscope(14)** Program-EEEP(3) Program-FastHN(1) Program-FlatNJ(1) Program-Fylogenetica(2) Program-GalledTree(1) Program-GraphDTL(2) Program-HapBound(1) Program-HGT_simul(1) Program-HiDe(1) Program-HorizStory(2) Program-Hybrid-Lambda(1) Program-HybridInterleave(5) Program-HybridNET(2) Program-HybridNumber(3) Program-Hybroscale(3) Program-JML(1) Program-LatTrans(5) Program-LEV1ATHAN(1) Program-Level1Generator(1) Program-Level2(2) Program-lingpy(1) Program-LNetwork(2) Program-MaafB(2) Program-Marlon(3) Program-MC-Net(1) Program-McKITsch(1) Program-Mowgli(5) Program-MowgliNNI(2) Program-MPNet(1) Program-MY-CLOSURE(1) Program-Nepal(7) Program-NetGen(3) Program-NetTest(1) Program-NetView(1) Program-Network(5) Program-Notung(1) Program-PADRE(7) Program-Phangorn(2) Program-PhippsNetwork(2) **Program-PhyloNet(11)** Program-PhyloNet-HMM(1) Program-PIRN(3) Program-Prunier(2) Program-Pyramids(3) Program-QNet(4) Program-Quartet(1) Program-Quartet-Decomposition(1) Program-QuartetMethods(1) Program-QuartetNet(2) Program-QuasiDec(1) Program-QuickCass(1) Program-RANGER-DTL(4) Program-RecMin(1) Program-Recodon(3) Program-RecPars(1) Program-Reticlad(2) Program-SAGE(2) Program-SAQ-Net(1) Program-Serial-NetEvolve(1) Program-SHRUB(3) Program-Simplistic(3) Program-Sliding-MinPD(1) Program-SNSA(2) Program-Spectronet(4) **Program-SplitsTree(36)** Program-SPNet(5) Program-SPRDist(1) Program-SuperQ(1) **Program-T-REX(13)** Program-TCS(8) Program-TERA(2) Program-TerminusEst(2) Program-TreeFix-DTL(1) Program-TreeMix(1) Program-Treevolve(2) Program-TripNet(2) Program-ultra-Net(1) Program-Ultratnet(1) Program-WeakHierarchies(2) Program-Xscape(1) **pyramid(8)** **quasi-median-network(3)** realization(4) recombination(29) recombination-detection(4) **reconstruction(255)** **regular-network(7)** reticulogram(10) serial-evolutionary-networks(1) simulated-annealing(4) simulation(5) site-consistency(1) **software(66)** split(27) split-decomposition(13) split-network(54) **SPR-distance(13)** **spread(2)** statistical-model(32) statistical-parsimony(3) supernet(4) survey(31) tanglegram(1) **time-consistent-network(12)** **tree-child-network(15)** **tree-sibling-network(10)** tripartition-distance(9) triplet-distance(3) unicyclic-network(3) **visualization(30)** weak-hierarchy(8) weakly-compatible(3)

input

software

classes

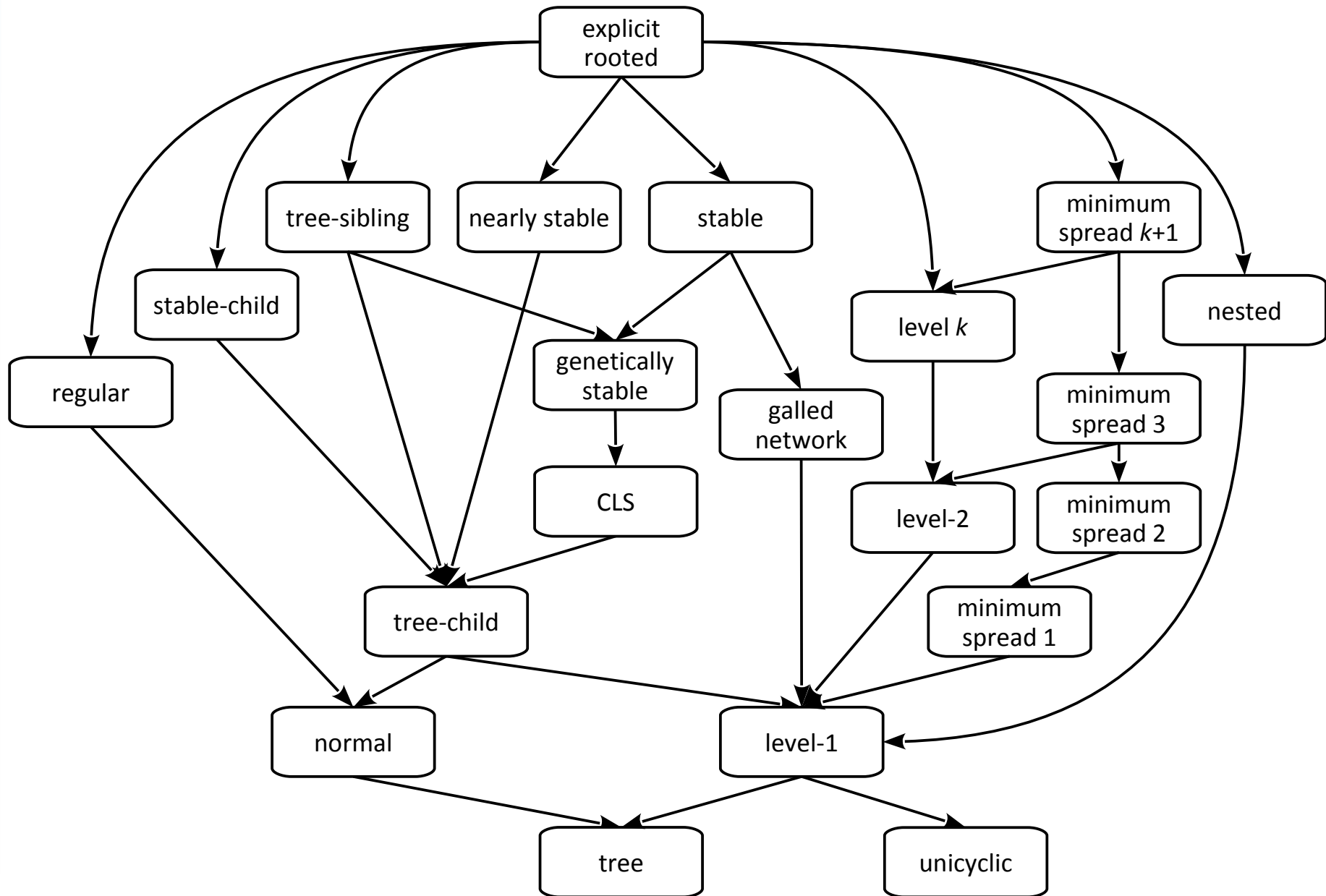
problems

algorithmic properties

Outline

- Phylogenetic networks
- Who is who in phylogenetic networks?
- **Classes of phylogenetic networks**
- The Tree Containment Problem
- A quadratic-time algorithm on quasi-stable networks

Classes of Phylogenetic Networks



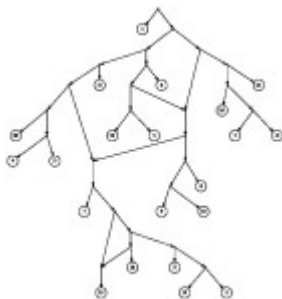
Outline

- Phylogenetic networks
- Who is who in phylogenetic networks?
- Classes of phylogenetic networks
- **The Tree Containment Problem**
- A quadratic-time algorithm on quasi-stable networks

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*

parsimony methods

*Hein 1990 - Kececioglu & Gusfield 1994 - Jin, Nakhleh,
Snir, Tuller 2009 - Park, Jin & Nakhleh 2010 - Kannan &
Wheeler, 2012*

likelihood methods

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

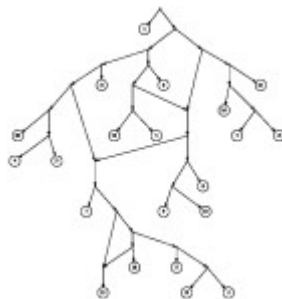
network *N*

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*

parsimony methods

*Hein 1990 - Kececioglu & Gusfield 1994 - Jin, Nakhleh,
Snir, Tuller 2009 - Park, Jin & Nakhleh 2010 - Kannan &
Wheeler, 2012*

likelihood methods

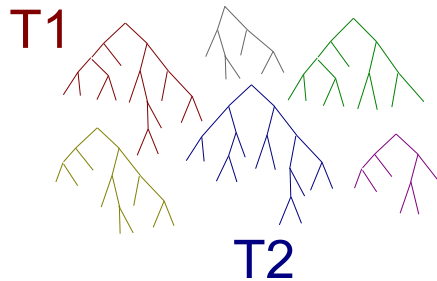
*Snir & Tuller 2009 - Jin, Nakhleh, Snir, Tuller 2009 -
Velasco & Sober 2009 - Meng & Kubatko 2009*

network *N*

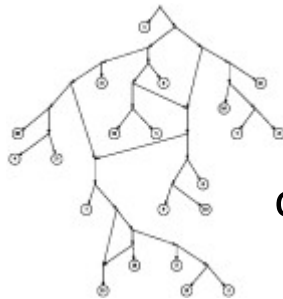
Phylogenetic network reconstruction

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

G1 G2



explicit network



{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

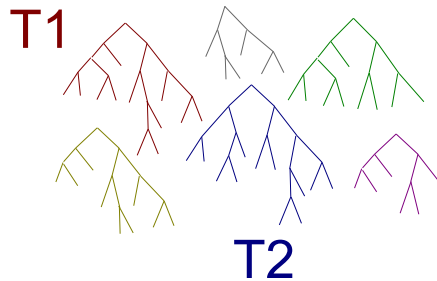
optimal super-network N:

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

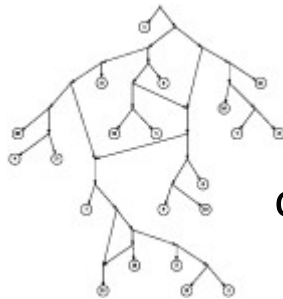
Phylogenetic network reconstruction

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

G1 G2



explicit network



{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

1470 species, >290 000 trees



Tree reconciliation or consensus

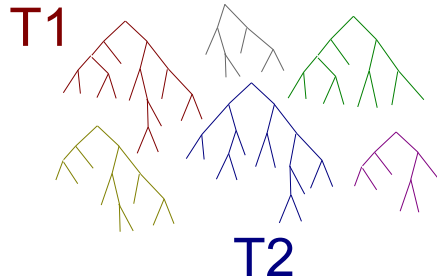
optimal super-network N:

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

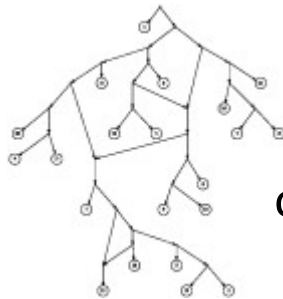
Phylogenetic network reconstruction

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



explicit network



{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

1470 species, >290 000 trees

Tree reconciliation or consensus

Tree Containment Problem

optimal super-network N:

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

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

Input: A binary phylogenetic network N and a tree T over the same set of taxa.

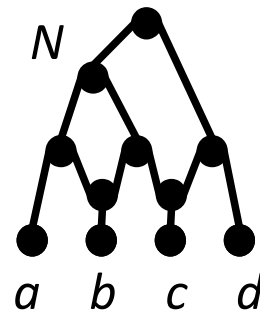
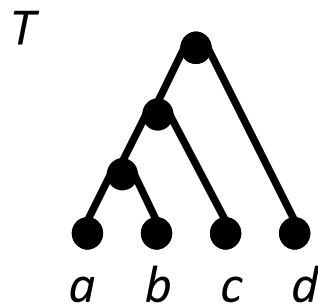
Question: Does N display T ?

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 ?

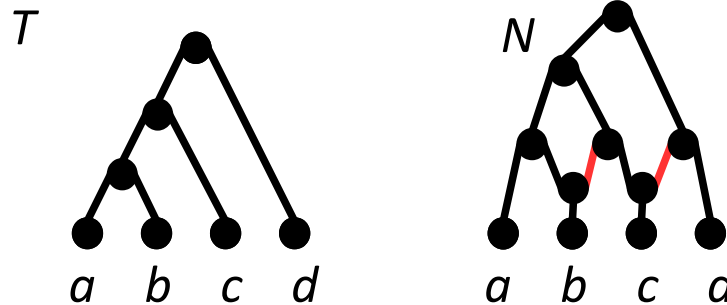


The Tree Containment Problem (T.C.P.)

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.)

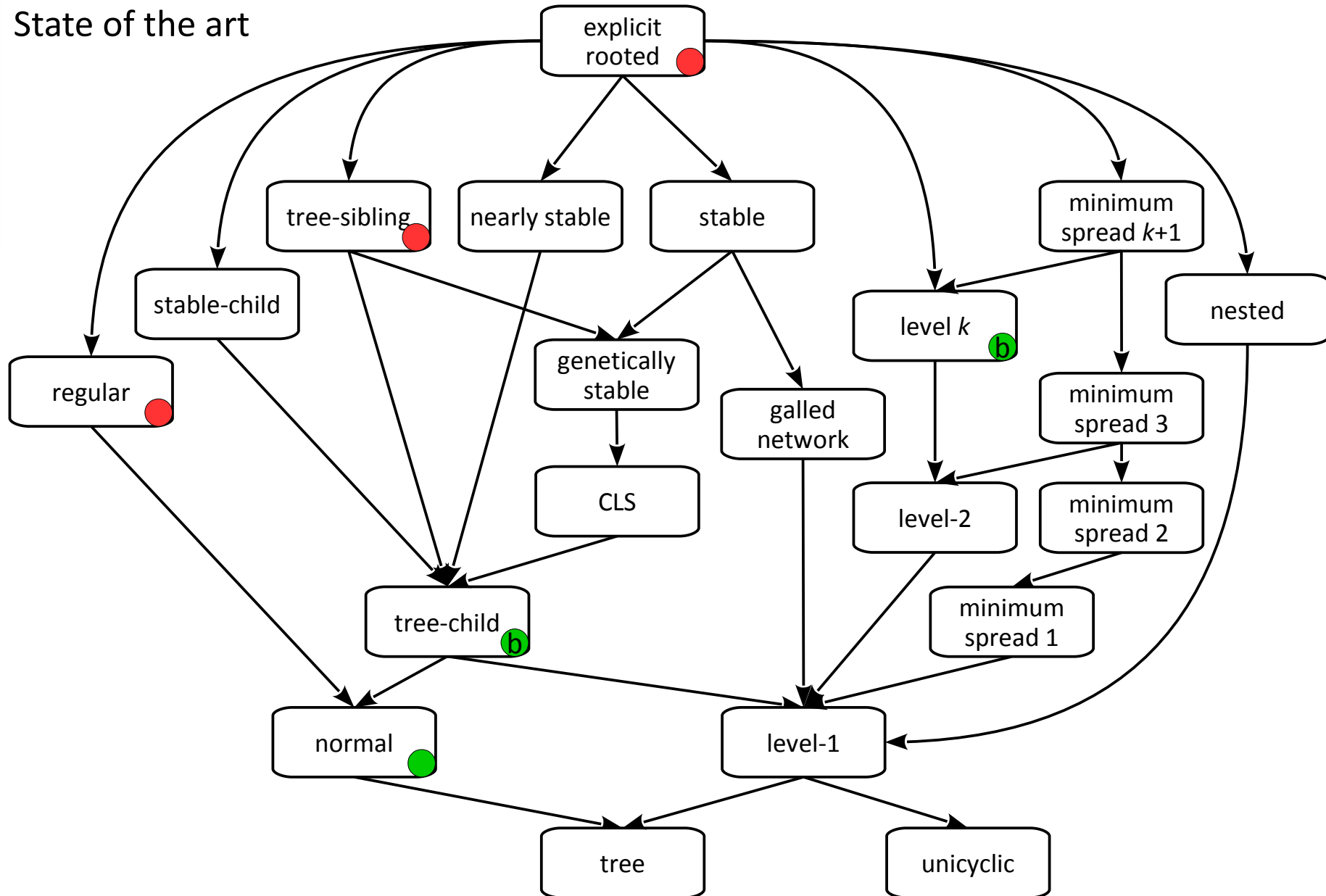
Input: A binary phylogenetic network N and a tree T over the same set of taxa.

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)

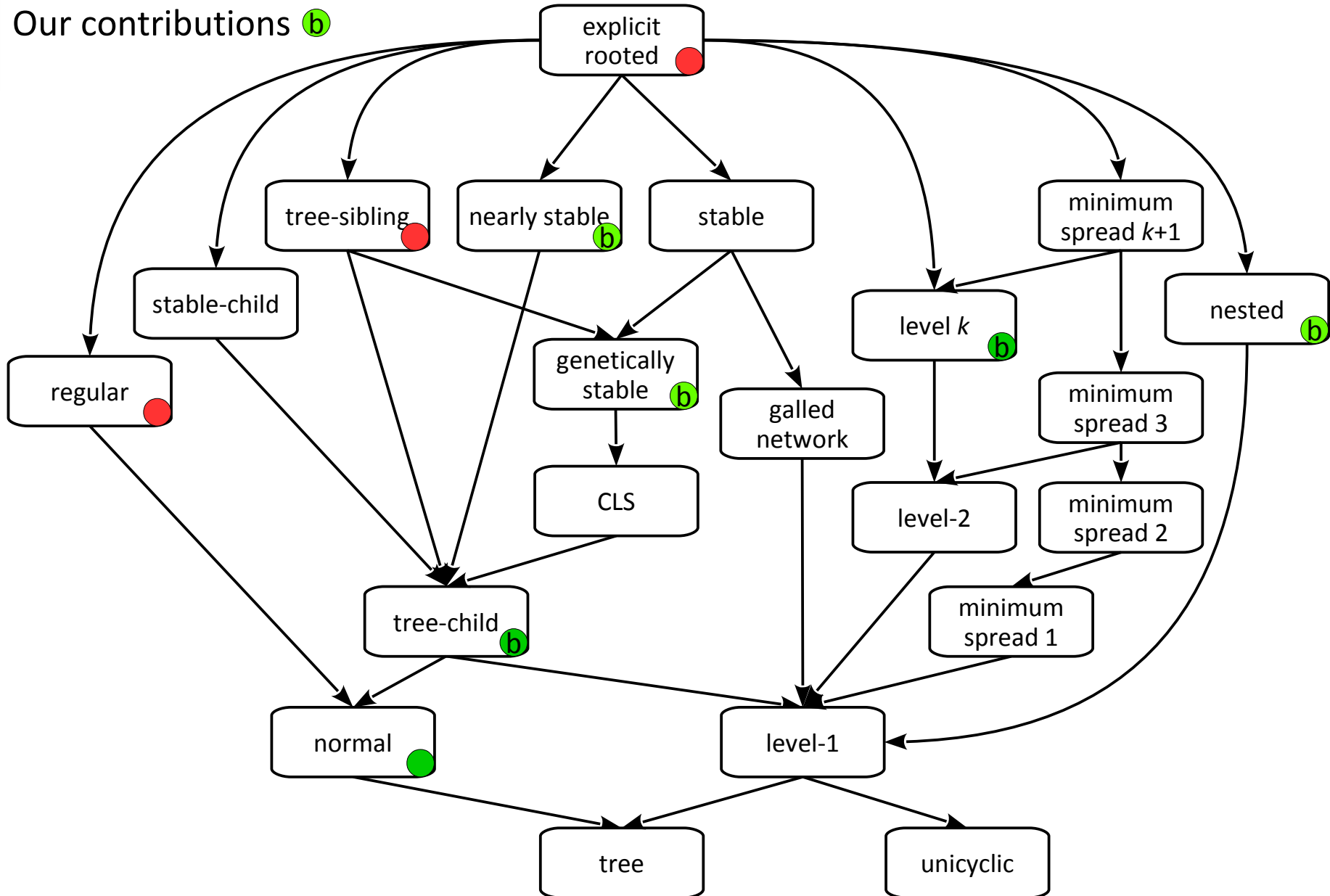
Classes of Phylogenetic Networks and the T.C.P.

State of the art



Classes of Phylogenetic Networks and the T.C.P.

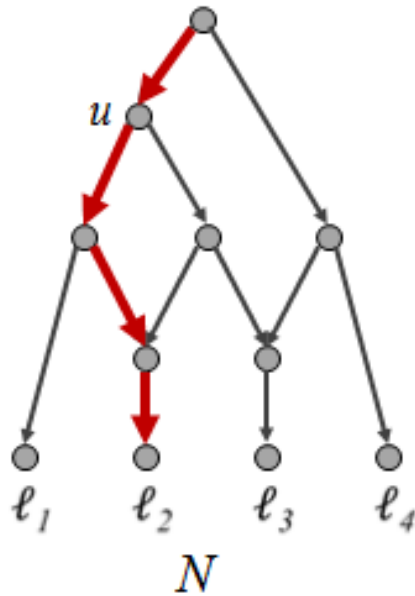
Our contributions b



Outline

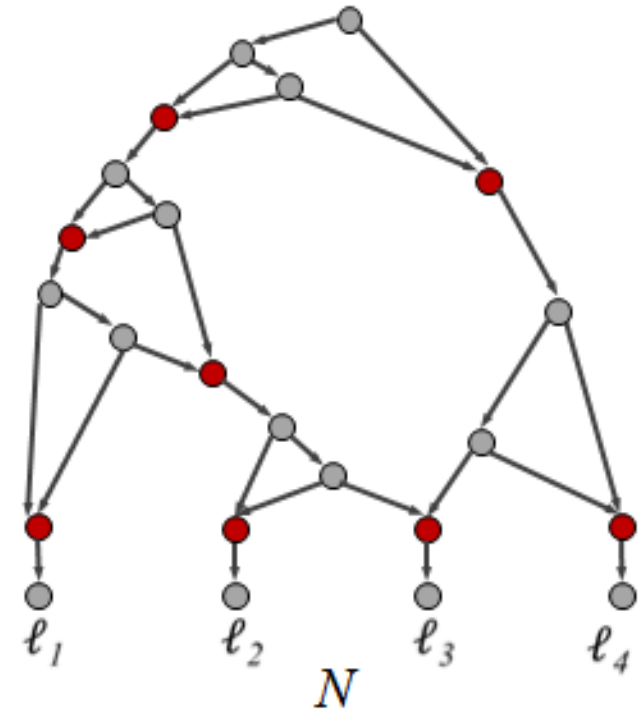
- Phylogenetic networks
- Who is who in phylogenetic networks?
- Classes of phylogenetic networks
- The Tree Containment Problem
- A quadratic-time algorithm on quasi-stable networks

Stable 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 **stable** if every reticulation vertex is stable.



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

Strategy to get a quadratic time algorithm for T.C.P.

Given N , a phylogenetic network with n leaves and the input tree T of the T.C.P.

Theorem 1: If N is stable then: $\#\{\text{reticulation vertices of } N\} \leq 4(n-1)$
 $\#\{\text{vertices of } N\} \leq 9n$

Theorem 2: If N is nearly-stable then $\#\{\text{reticulation vertices of } N\} \leq 12(n-1)$

Theorem 3: Considering a longest path in N , and following a case analysis:

- either realize that T is not contained in N
- or build a network N' with less arcs than N such that T contained in N if and only if T contained in N'

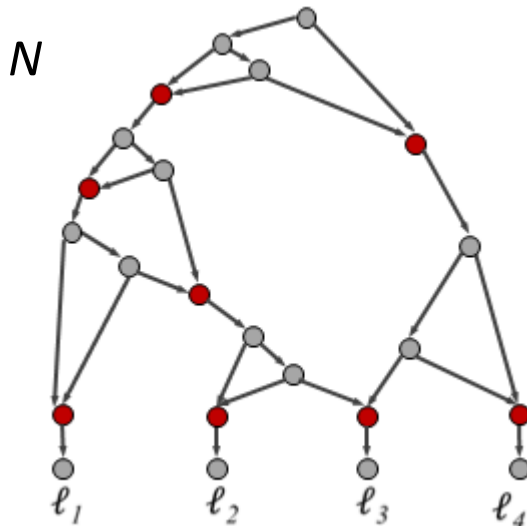
Number of reticulations of a stable network

Decompose N into $2n-2$ paths:

- remove one reticulation arc per reticulation, ensuring we get no « dummy leaf », to get a tree T with n leaves
- summarize T into a rooted binary tree T' with n leaves... and $2n-2$ arcs

We can prove (technical) that:

each path contains at most 2 reticulation vertices



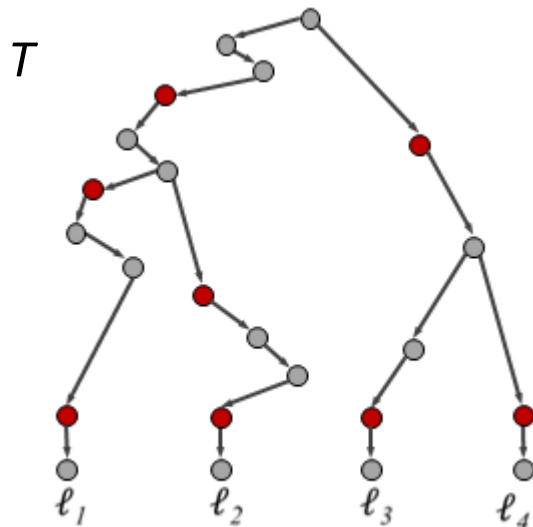
Number of reticulations of a stable network

Decompose N into $2n-2$ paths:

- remove one reticulation arc per reticulation, ensuring we get no « dummy leaf », to get a tree T with n leaves
- summarize T into a rooted binary tree T' with n leaves... and $2n-2$ arcs

We can prove (technical) that:

each path contains at most 2 reticulation vertices



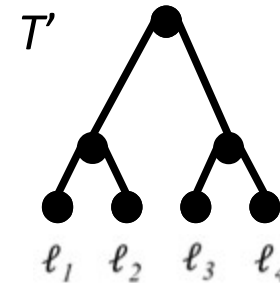
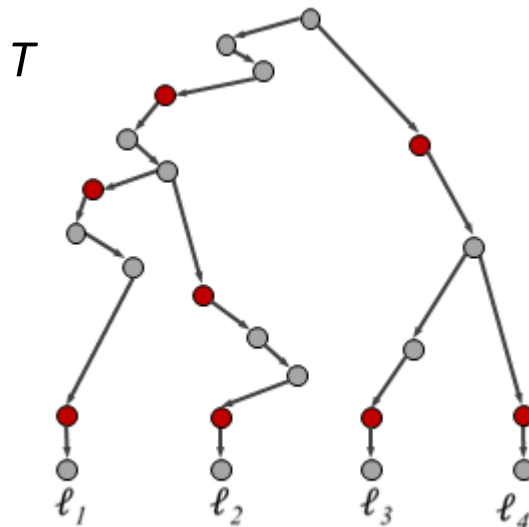
Number of reticulations of a stable network

Decompose N into $2n-2$ paths:

- remove one reticulation arc per reticulation, ensuring we get no « dummy leaf », to get a tree T with n leaves
- summarize T into a rooted binary tree T' with n leaves... and $2n-2$ arcs

We can prove (technical) that:

each path contains at most 2 reticulation vertices



Number of reticulations of a stable network

Decompose N into $2n-2$ paths:

- remove one reticulation arc per reticulation, ensuring we get no « dummy leaf », to get a tree T with n leaves
- summarize T into a rooted binary tree T' with n leaves... and $2n-2$ arcs

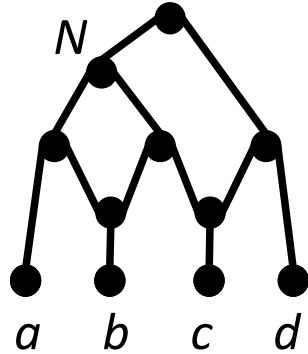
We can prove (technical) that:

each path contains at most 2 reticulation vertices

→ N contains at most $4(n-1)$ reticulation vertices

Number of reticulations of a stable network

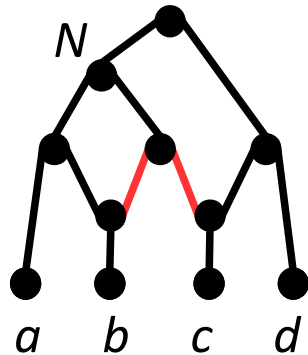
« Dummy leaves »?



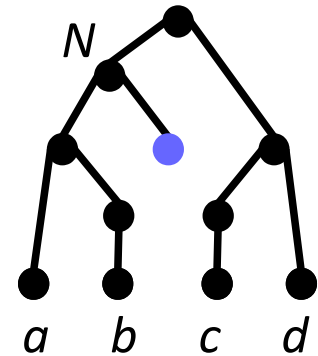
Deleting reticulation arcs can create « **dummy leaves** »

Number of reticulations of a stable network

« Dummy leaves »?



Deleting **reticulation arcs** can create « **dummy leaves** »



Number of reticulations of a stable network

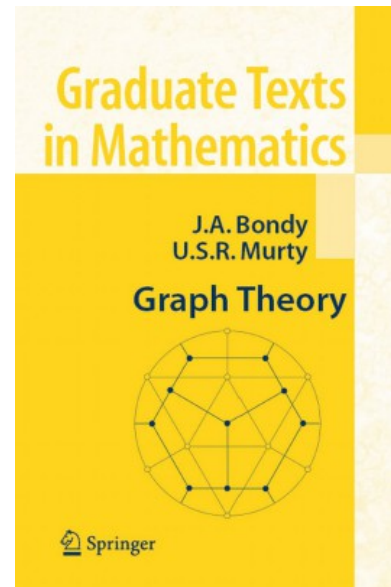
Possible to avoid creating « dummy leaves »?

Number of reticulations of a stable network

Possible to avoid creating « dummy leaves »?

16.2 Matchings in Bipartite Graphs 423

16.2.13 Let $G := G[X, Y]$ be a bipartite graph such that $d(x) \geq 1$ for all $x \in X$ and $d(x) \geq d(y)$ for all $xy \in E$, where $x \in X$ and $y \in Y$. Show that G has a matching covering every vertex of X .
(N. ALON)



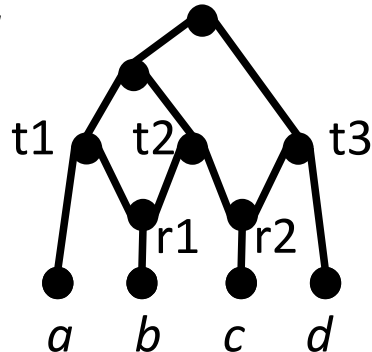
Number of reticulations of a stable network

Possible to avoid creating « dummy leaves »?

16.2 Matchings in Bipartite Graphs 423

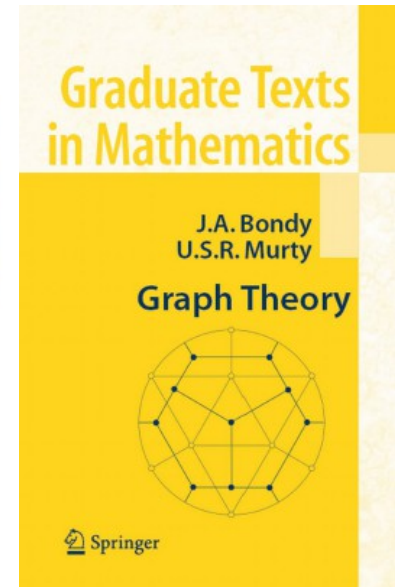
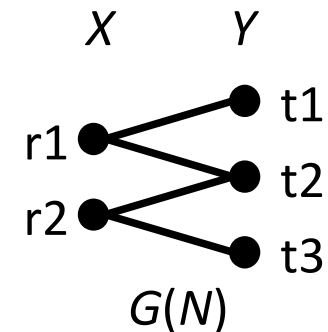
16.2.13 Let $G := G[X, Y]$ be a bipartite graph such that $d(x) \geq 1$ for all $x \in X$ and $d(x) \geq d(y)$ for all $xy \in E$, where $x \in X$ and $y \in Y$. Show that G has a matching covering every vertex of X . (N. ALON)

Given N



Build $G(N)$, bipartite graph such that:

- X = reticulation vertices of N
 → all vertices in X have degree 2
- Y = tree vertices of N with at least one reticulation child
 → all vertices in Y have degree 1 or 2
- edge between x and y iff x is a child of y

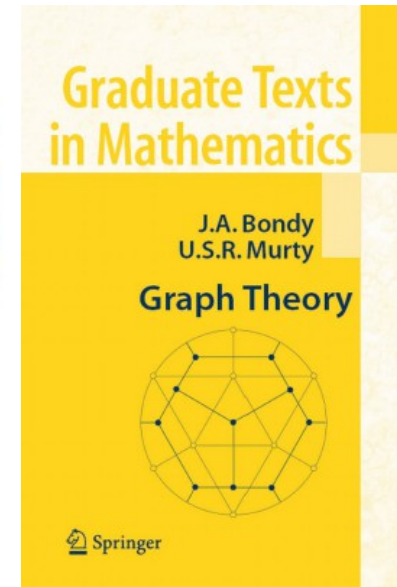


Number of reticulations of a stable network

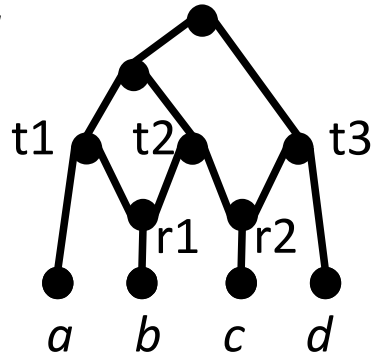
Possible to avoid creating « dummy leaves »?

16.2 Matchings in Bipartite Graphs 423

16.2.13 Let $G := G[X, Y]$ be a bipartite graph such that $d(x) \geq 1$ for all $x \in X$ and $d(x) \geq d(y)$ for all $xy \in E$, where $x \in X$ and $y \in Y$. Show that G has a matching covering every vertex of X . (N. ALON)

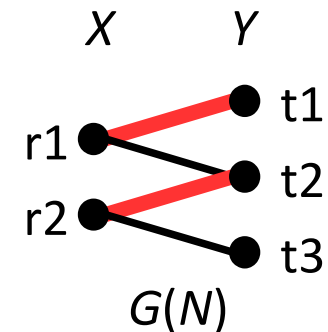


Given N



Build $G(N)$, bipartite graph such that:

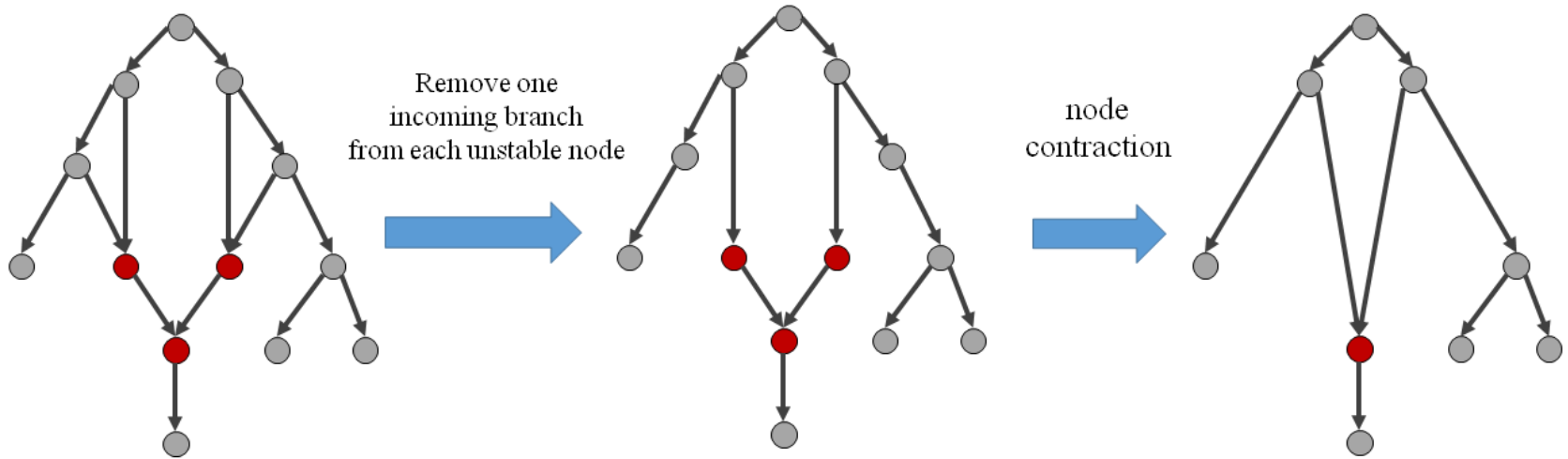
- X = reticulation vertices of N
 → all vertices in X have degree 2
- Y = tree vertices of N with at least one reticulation child
 → all vertices in Y have degree 1 or 2
- edge between x and y iff x is a child of y



→ **matching covering every vertex of X** → **edges to remove from N**

Number of reticulations of a nearly-stable network

Reduce nearly-stable networks to stable networks



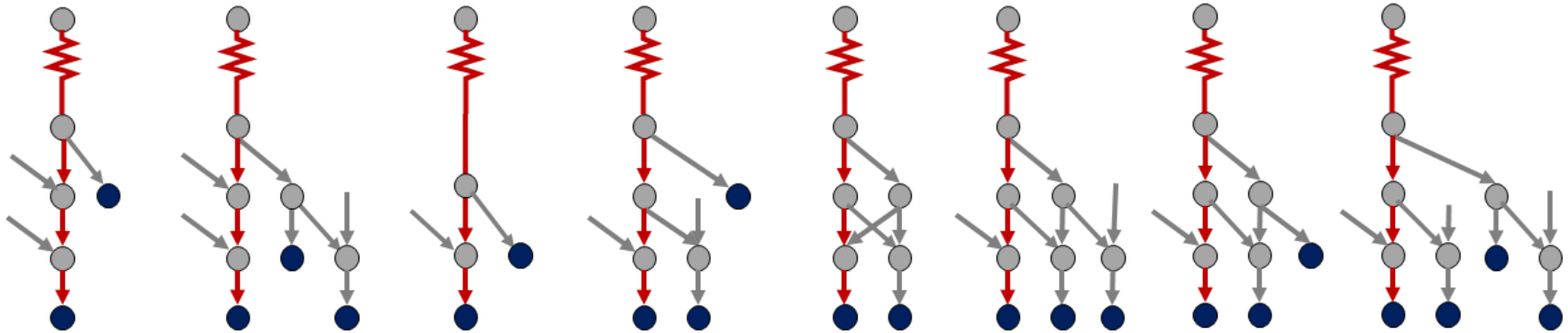
$$\#UnstableRet \leq 2 \quad \#StableRet \leq 8(n-1)$$

$$\#StableRet \leq 4(n-1)$$

Deleting reticulation arcs to simplify the question

Simplify N by removing an edge near the end of a **longest path P** .

Case analysis (8 cases):



Perspectives...

