ANR-FWF-MOST meeting
2018-10-30 - Wien

Phylogenetic networks: overview, subclasses and counting problems

Philippe Gambette

UNIVERSITÉ
PARIS-EST
MARNE-LA-VALLÉE

Gaspard - Monge
Laboratoire d'informatique

Outline

- An introduction to phylogenetic networks
- Classes of phylogenetic networks
- Counting problems on phylogenetic networks

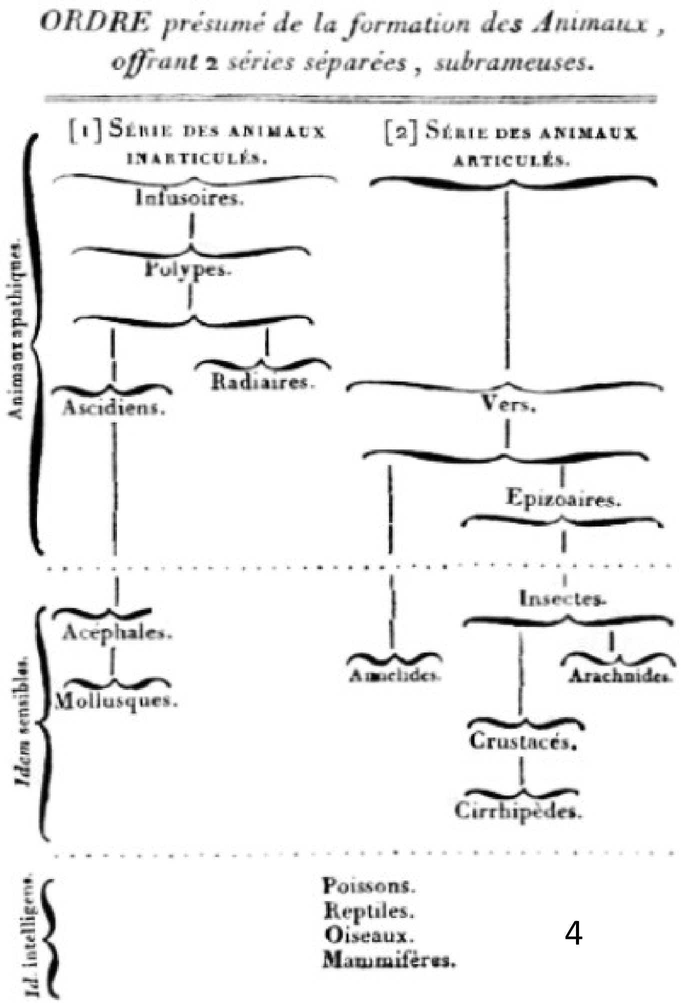
Outline

- An introduction to phylogenetic networks
- Classes of phylogenetic networks
- Counting problems on phylogenetic networks

Phylogenetic trees

Phylogenetic tree of a set of species:

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



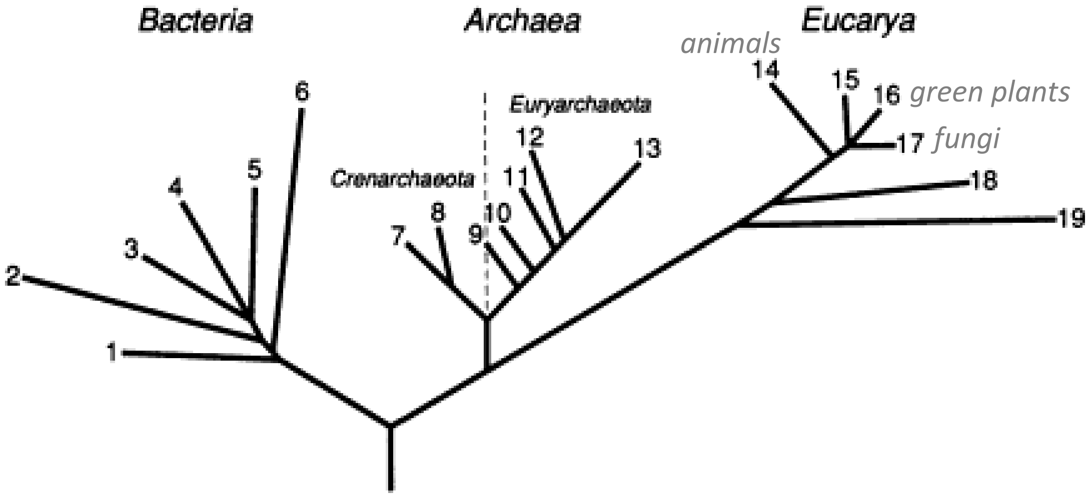
Lamarck : Histoire naturelle des animaux sans vertèbres (1815)

Phylogenetic trees

Phylogenetic tree of a set of species:

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

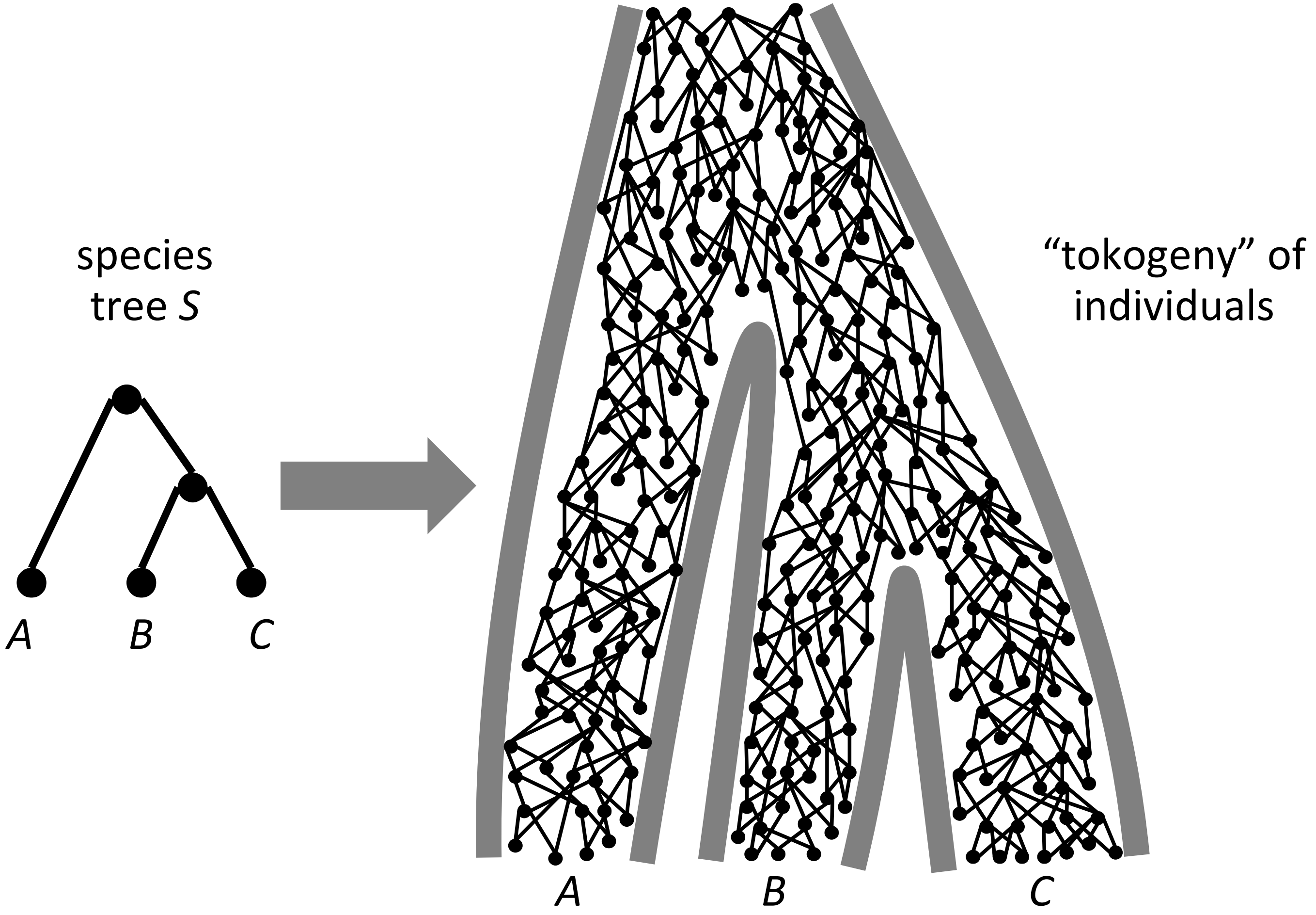
modelization



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

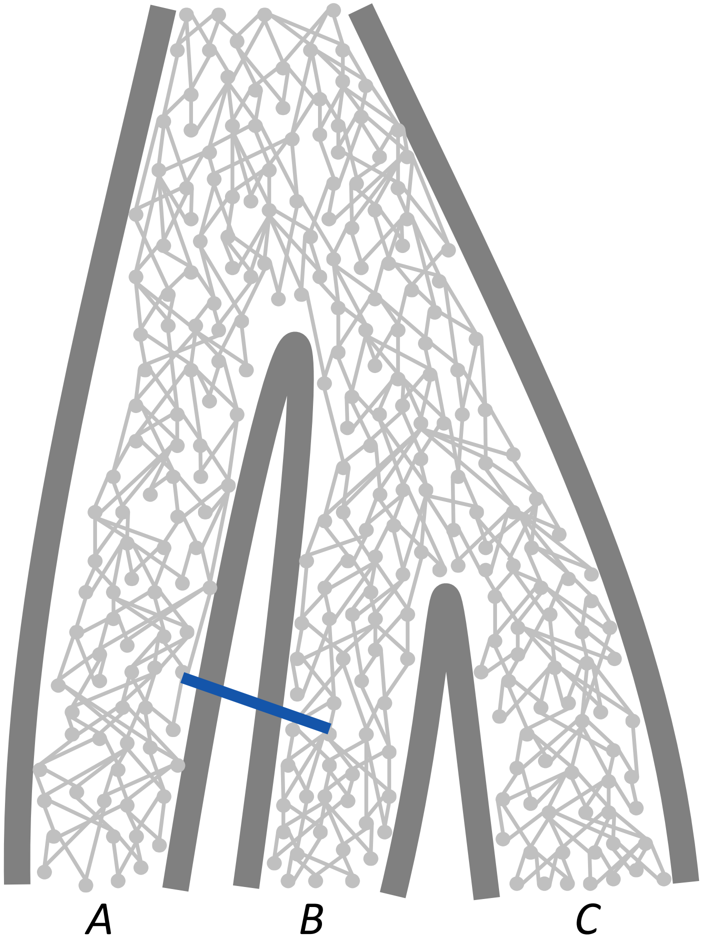
Phylogenetic tree of a set of species



Genetic material transfers

Transfers of genetic material between coexisting species:

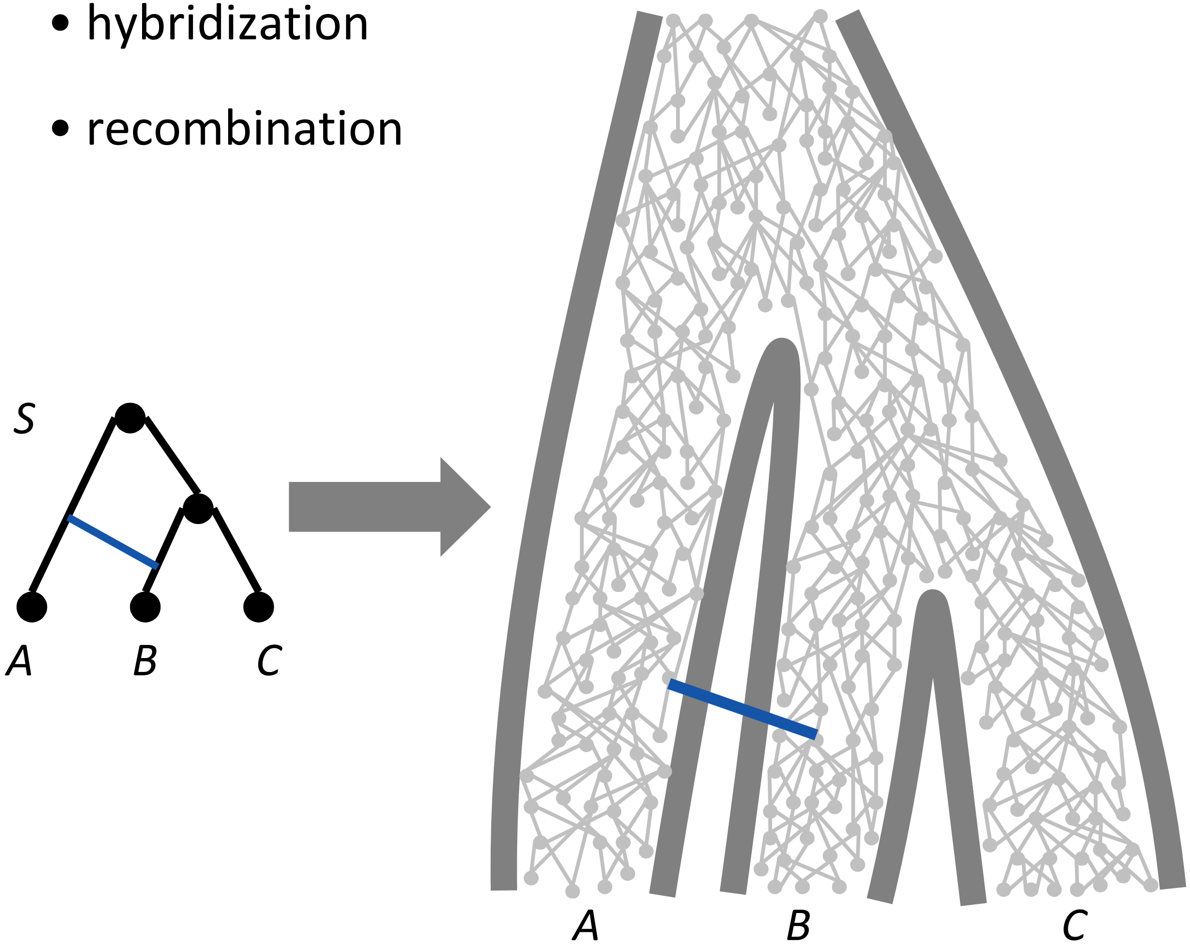
- lateral gene transfer
- hybridization
- recombination



Genetic material transfers

Transfers of genetic material between coexisting species:

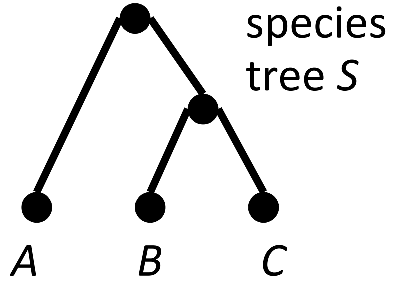
- lateral gene transfer
- hybridization
- recombination

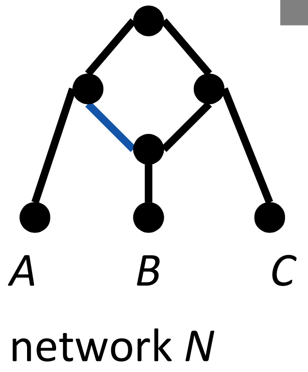


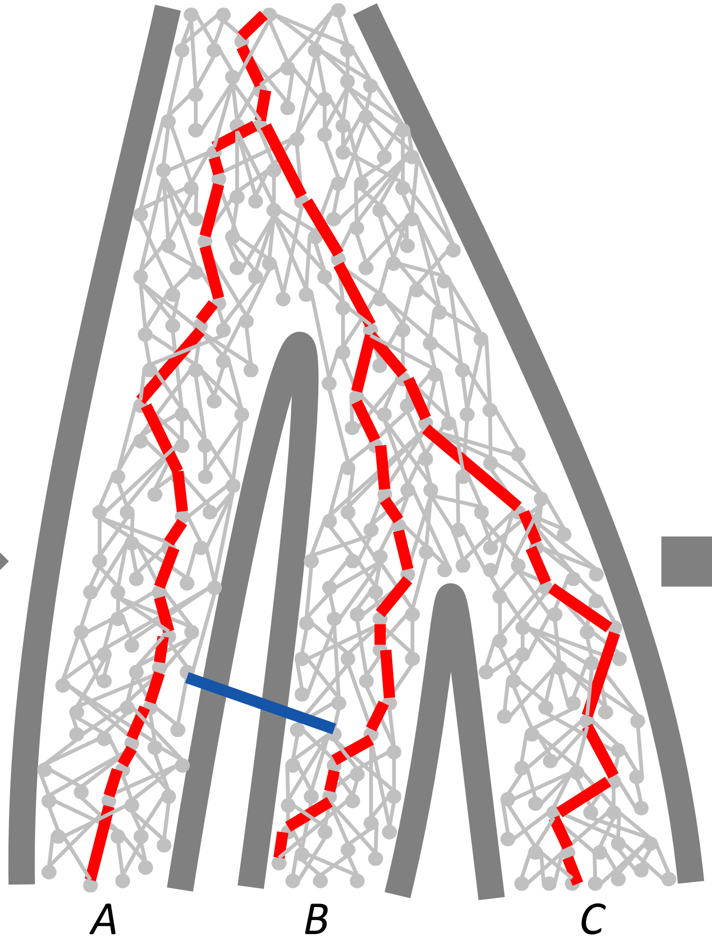
Genetic material transfers

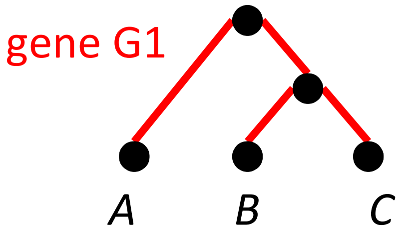
Transfers of genetic material between coexisting species:

- lateral gene transfer
- hybridization
- recombination





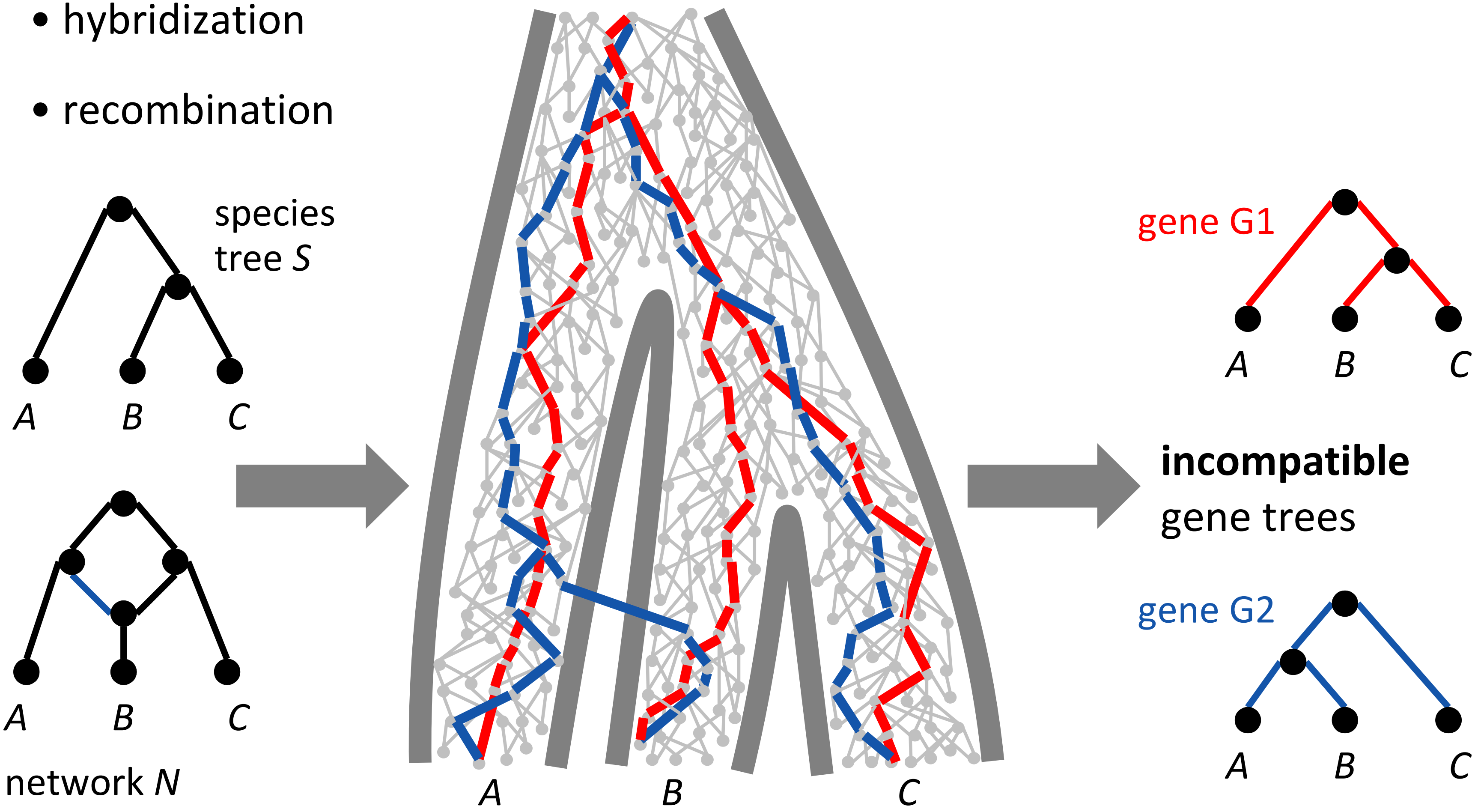




Genetic material transfers

Transfers of genetic material between coexisting species:

- lateral gene transfer
- hybridization
- recombination



species tree *S*

network *N*

gene G1

gene G2

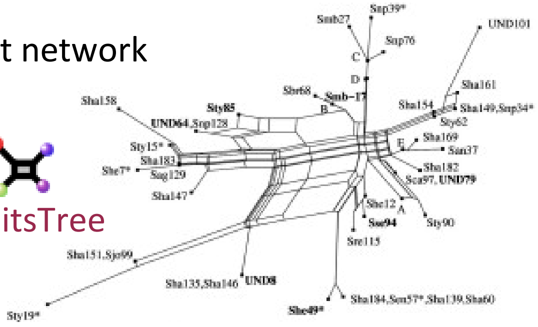
incompatible gene trees

Phylogenetic networks

Phylogenetic network: network representing evolution data

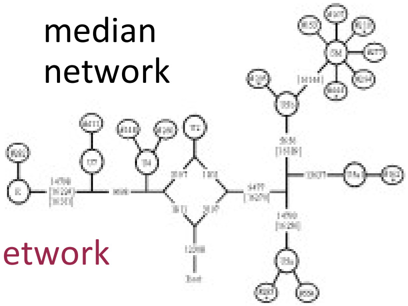
- **abstract / implicit / data-display** phylogenetic networks: to **classify**, **visualize** **data**

split network



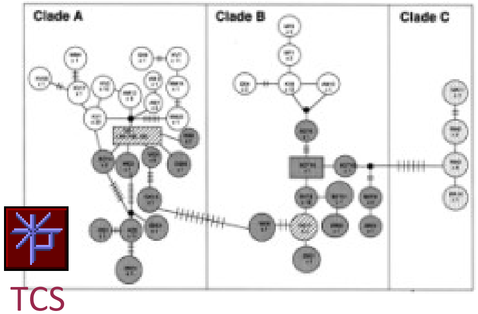
SplitsTree

median network



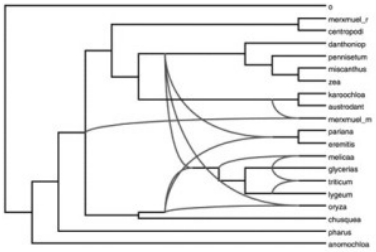
Network

minimum spanning network



TCS

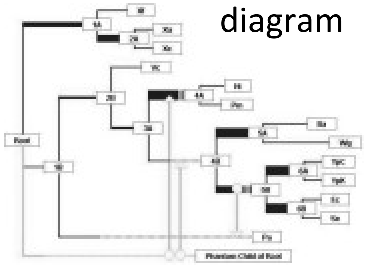
- **explicit** phylogenetic networks: to **model** **evolution**



galled network

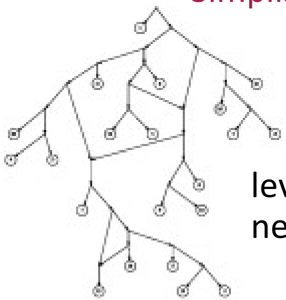
Dendroscope

synthesis diagram



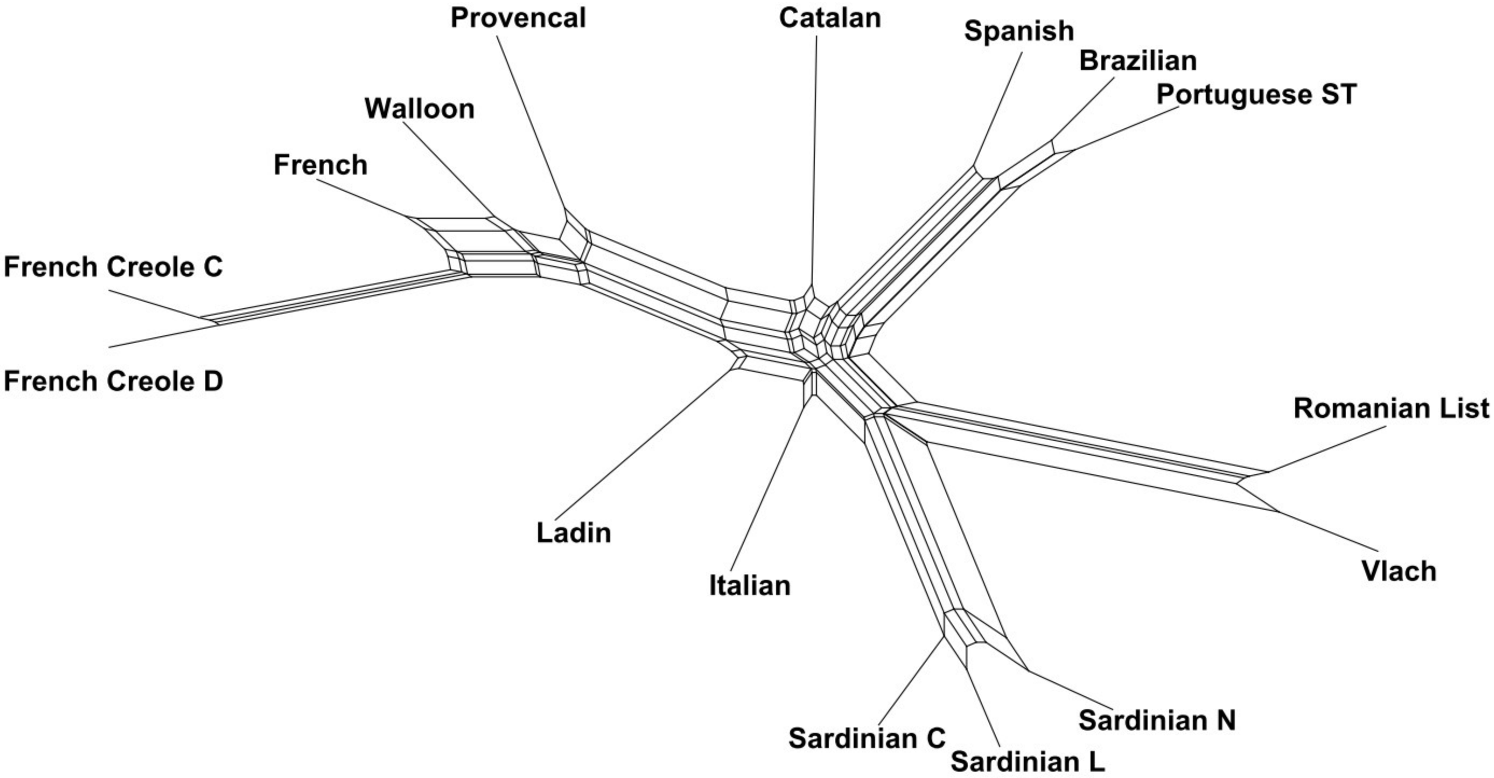
HorizStory

Simplistic



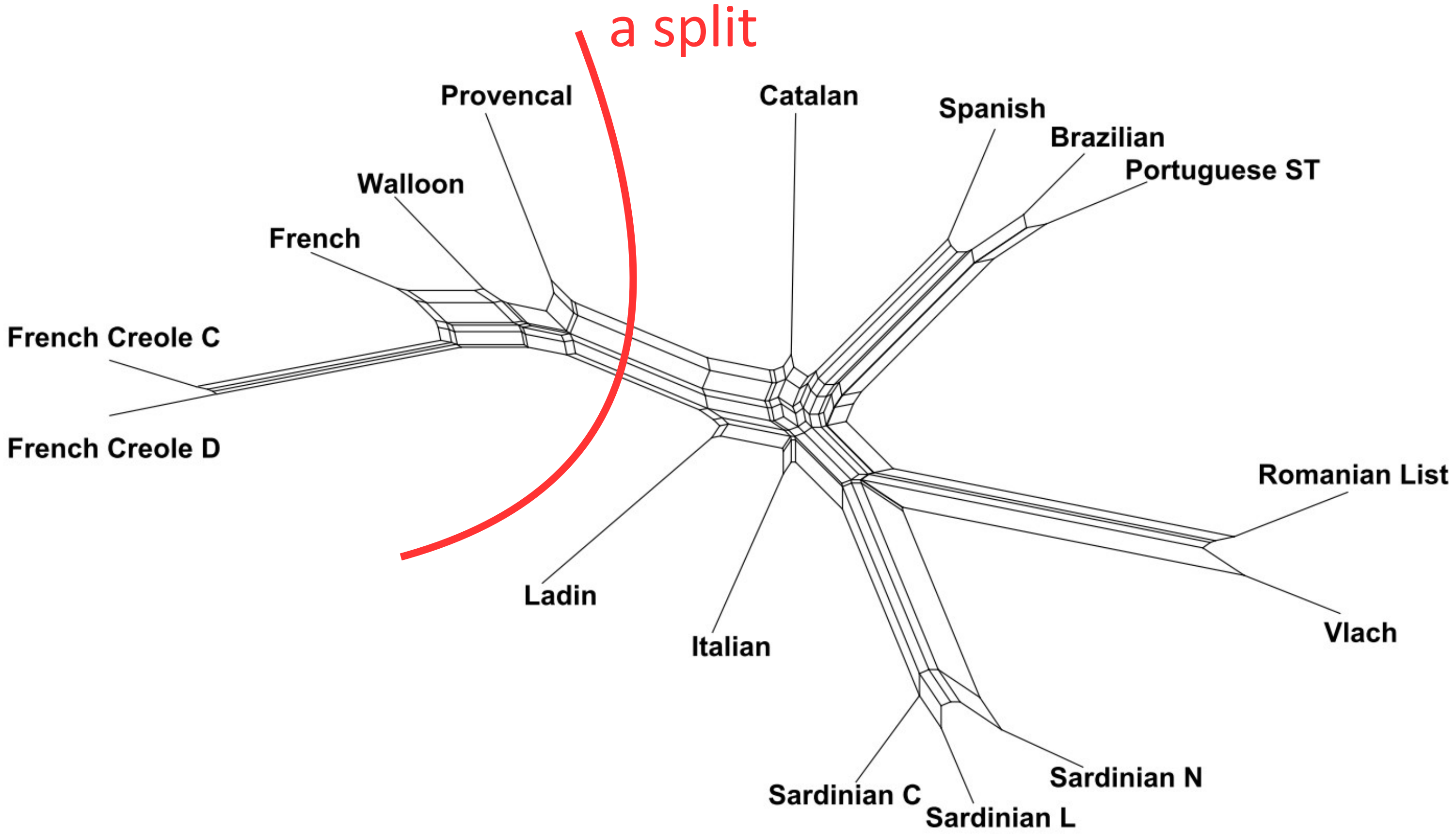
level-2 network

Abstract phylogenetic networks: split networks



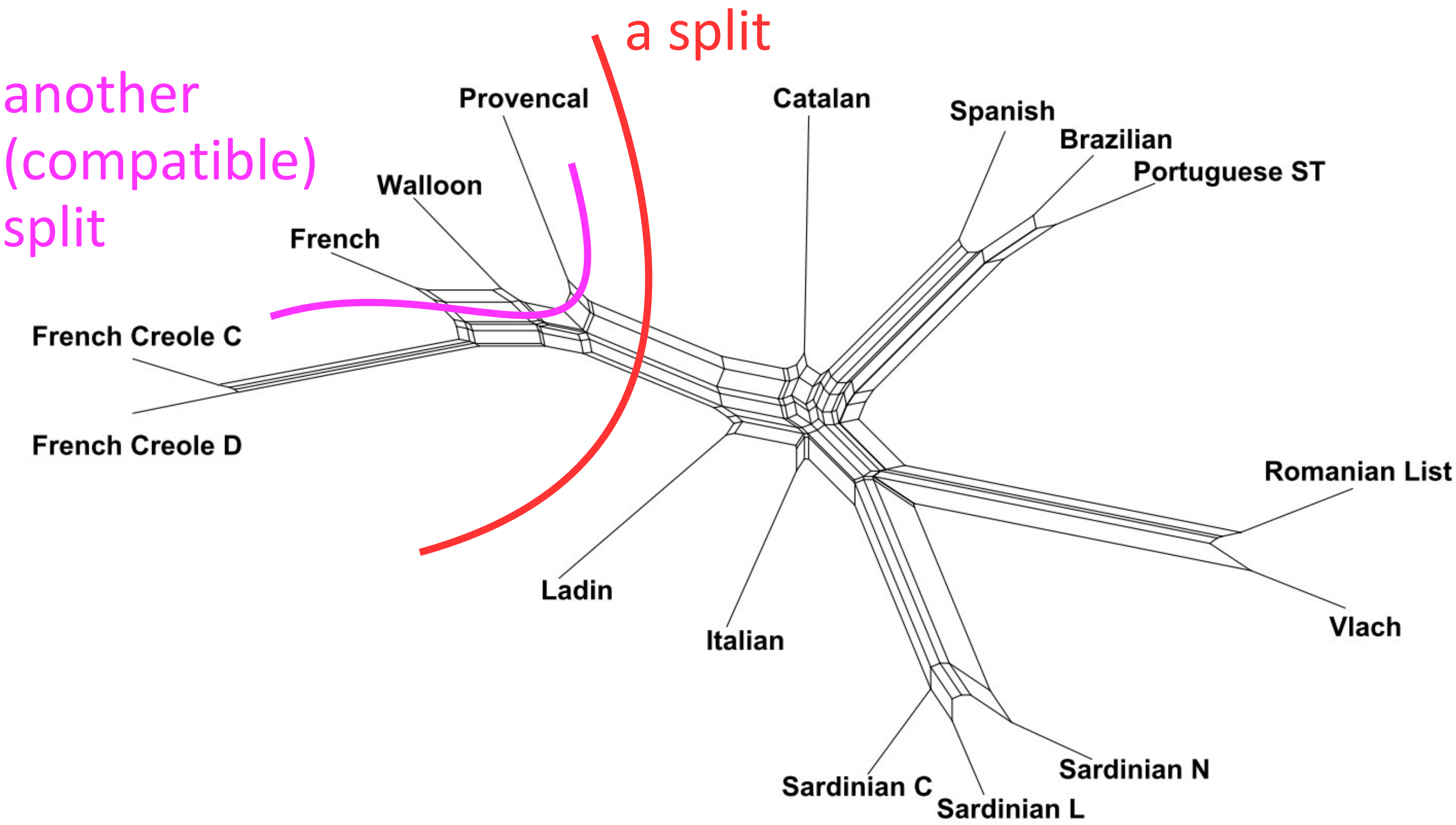
Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks



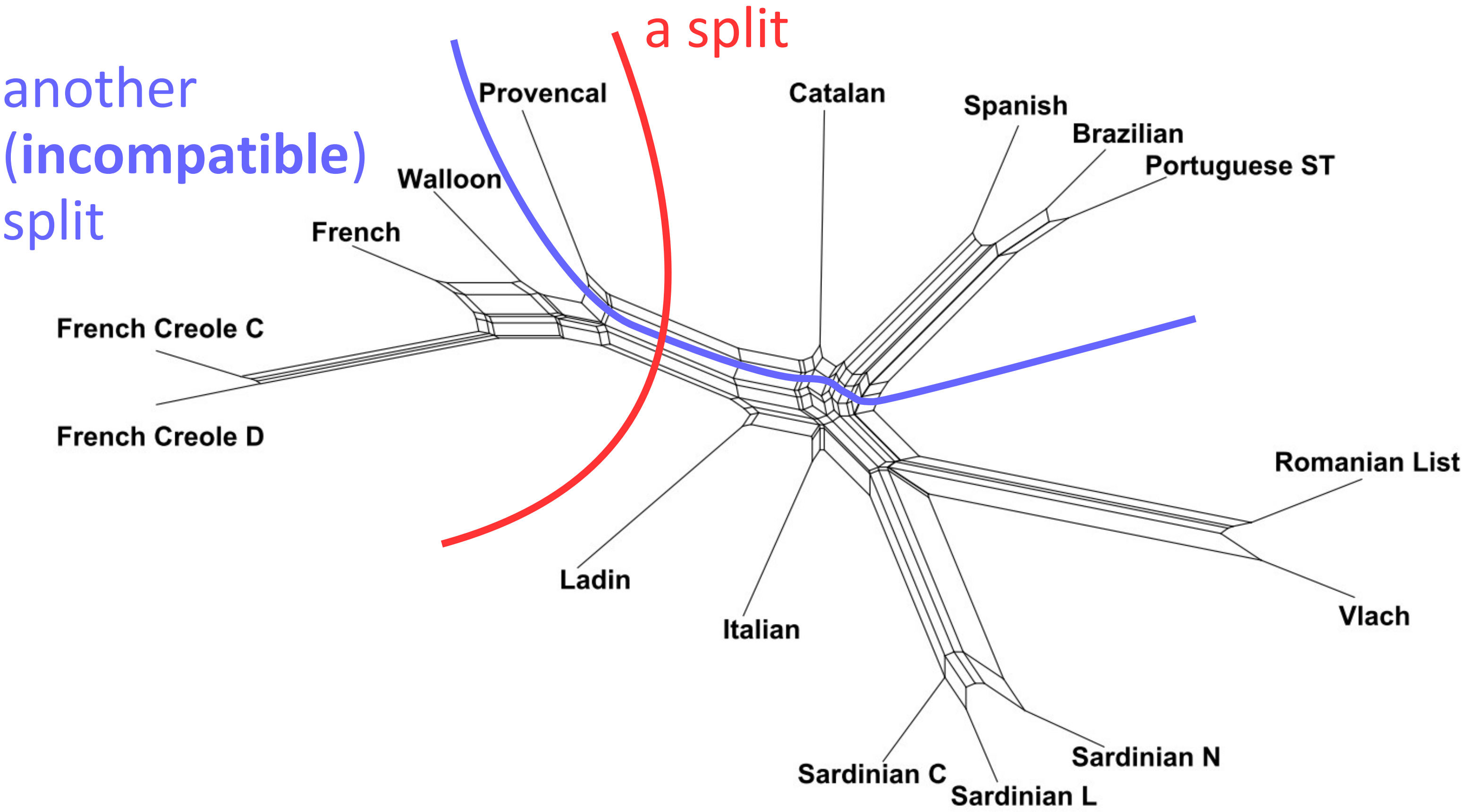
Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

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

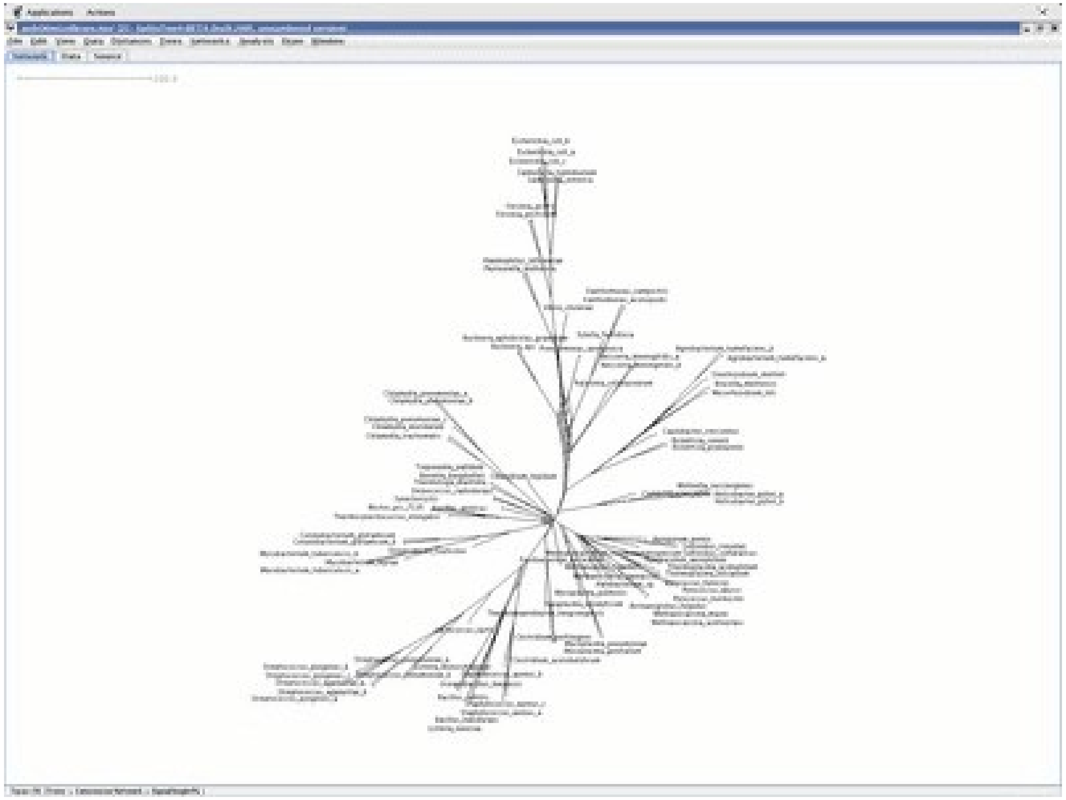


Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks

The “boxes” in split networks: incompatible splits

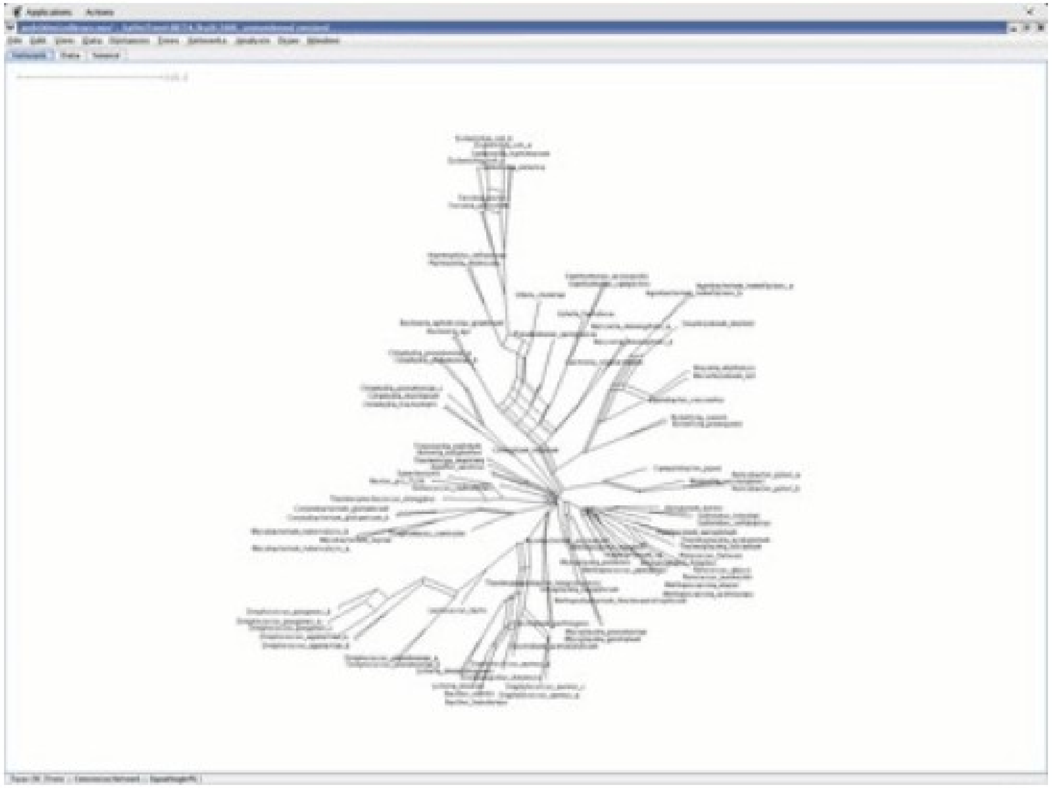
→ open them as much as possible in SplitsTree!



Abstract phylogenetic networks: split networks

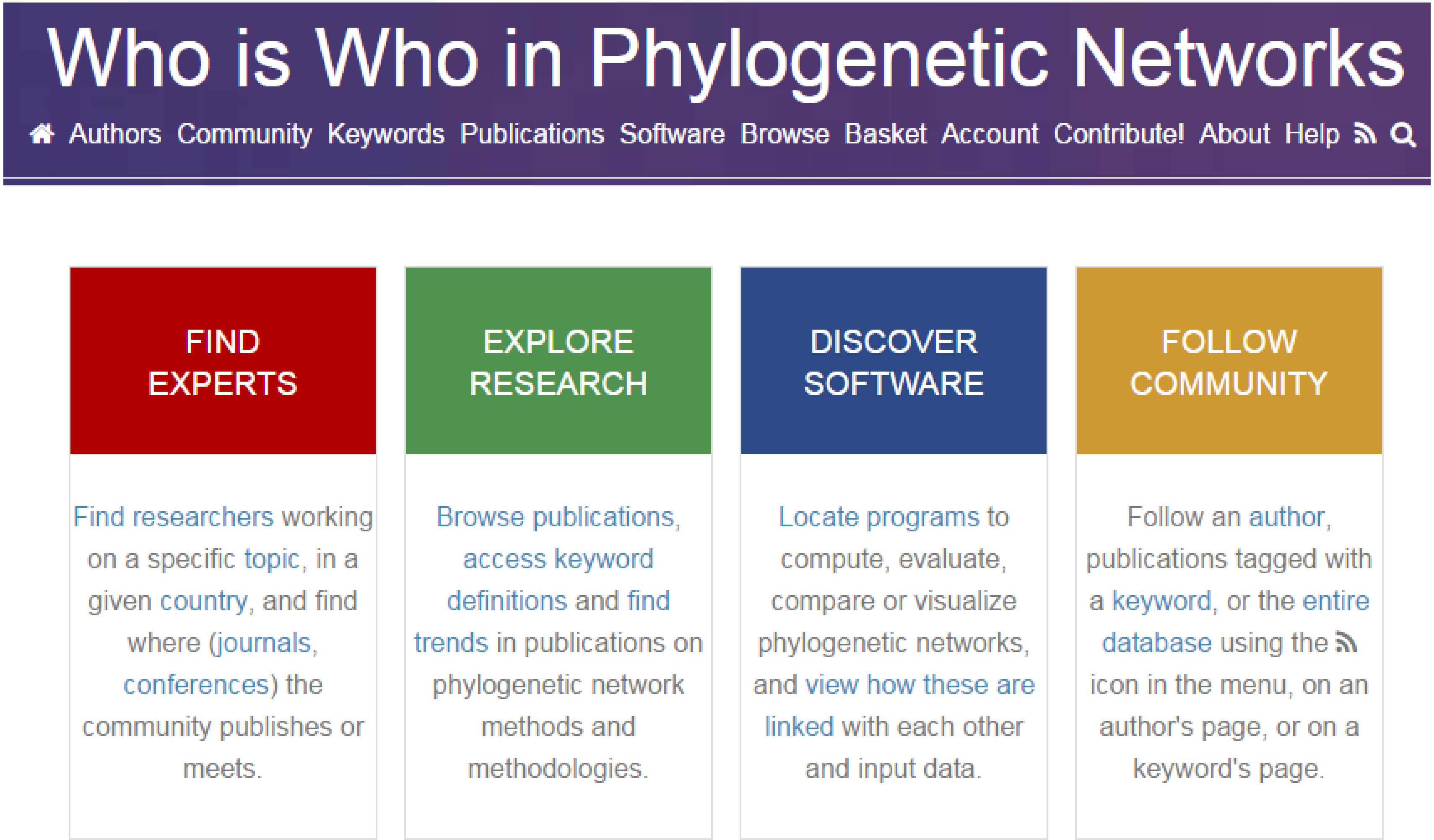
The “boxes” in split networks: incompatible splits

→ 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

Who is who in Phylogenetic Networks?



Who is Who in Phylogenetic Networks

FIND EXPERTS

Find researchers working on a specific topic, in a given country, and find where (journals, conferences) the community publishes or meets.

EXPLORE RESEARCH

Browse publications, access keyword definitions and find trends in publications on phylogenetic network methods and methodologies.

DISCOVER SOFTWARE

Locate programs to compute, evaluate, compare or visualize phylogenetic networks, and view how these are linked with each other and input data.

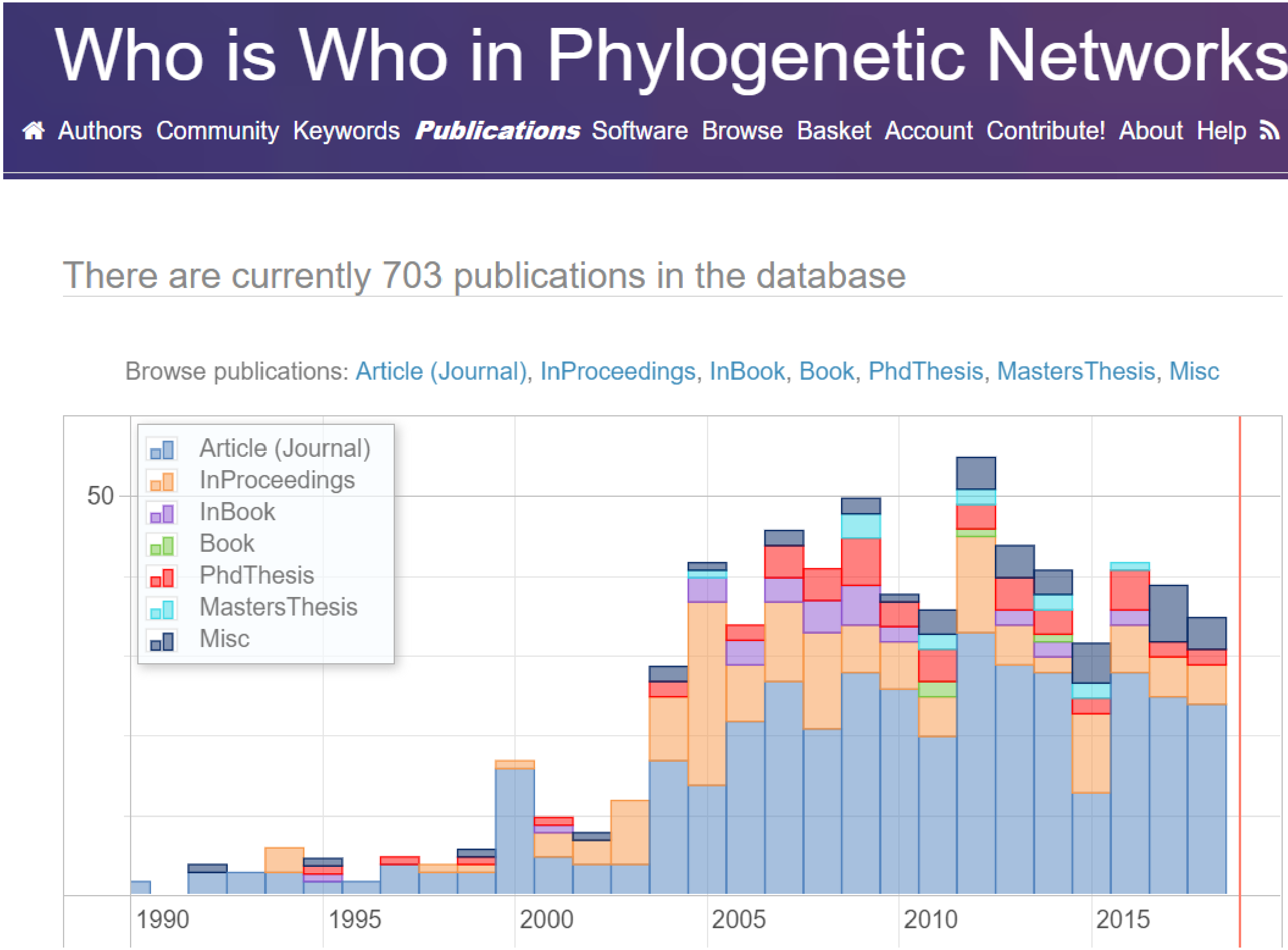
FOLLOW COMMUNITY

Follow an author, publications tagged with a keyword, or the entire database using the icon in the menu, on an author's page, or on a keyword's page.

Based on BibAdmin by Sergiu Chelcea
+ tag clouds, date histograms, journal lists, keyword definitions, co-author graphs (work of Tushar Agarwal)

http://phylnet.univ-mlv.fr

Who is who in Phylogenetic Networks?



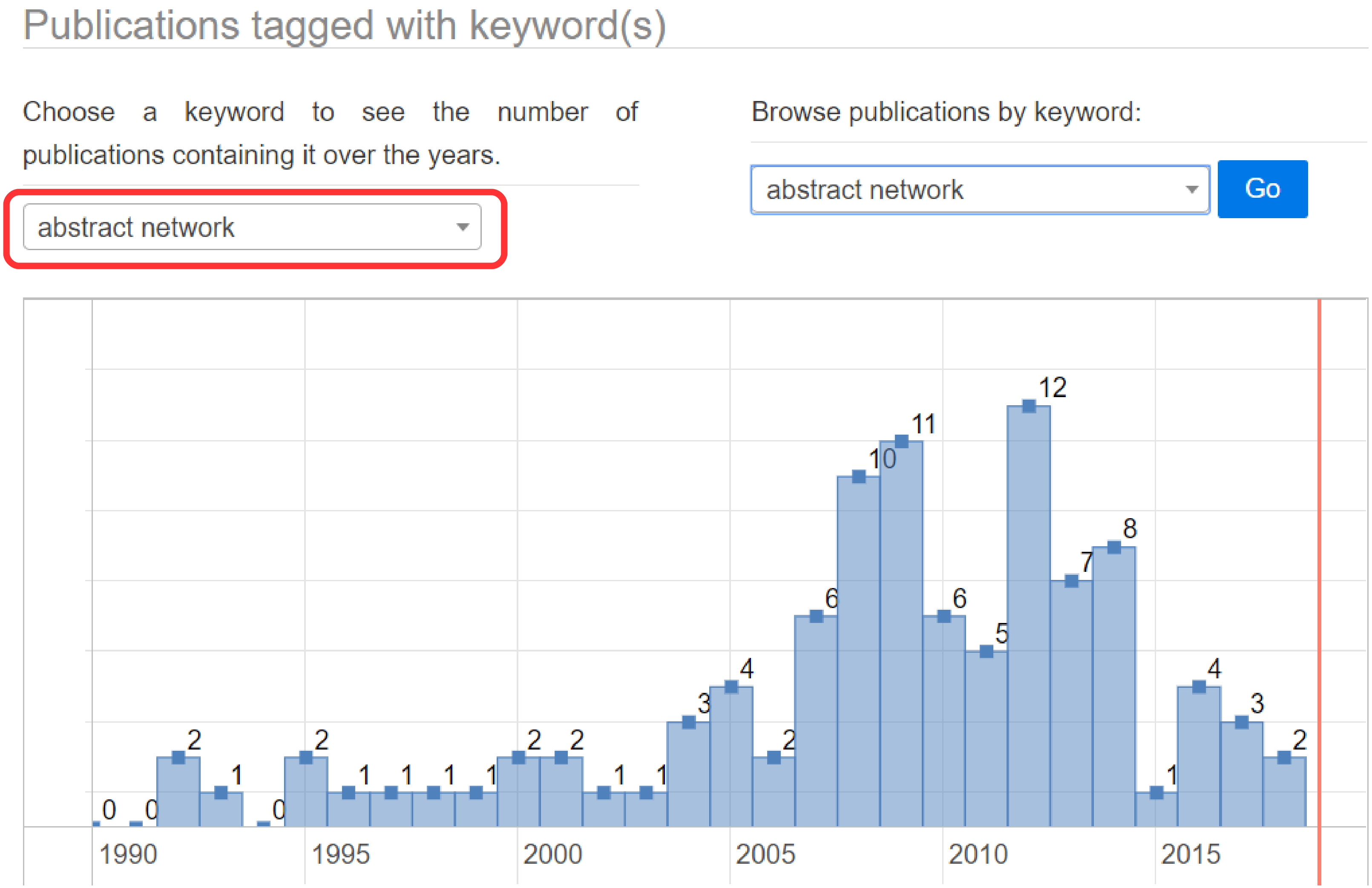
Who is Who in Phylogenetic Networks

There are currently 703 publications in the database

Browse publications: Article (Journal), InProceedings, InBook, Book, PhdThesis, MastersThesis, Misc

http://phylnet.univ-mlv.fr

Who is who in Phylogenetic Networks?



Publications tagged with keyword(s)

Choose a keyword to see the number of publications containing it over the years.

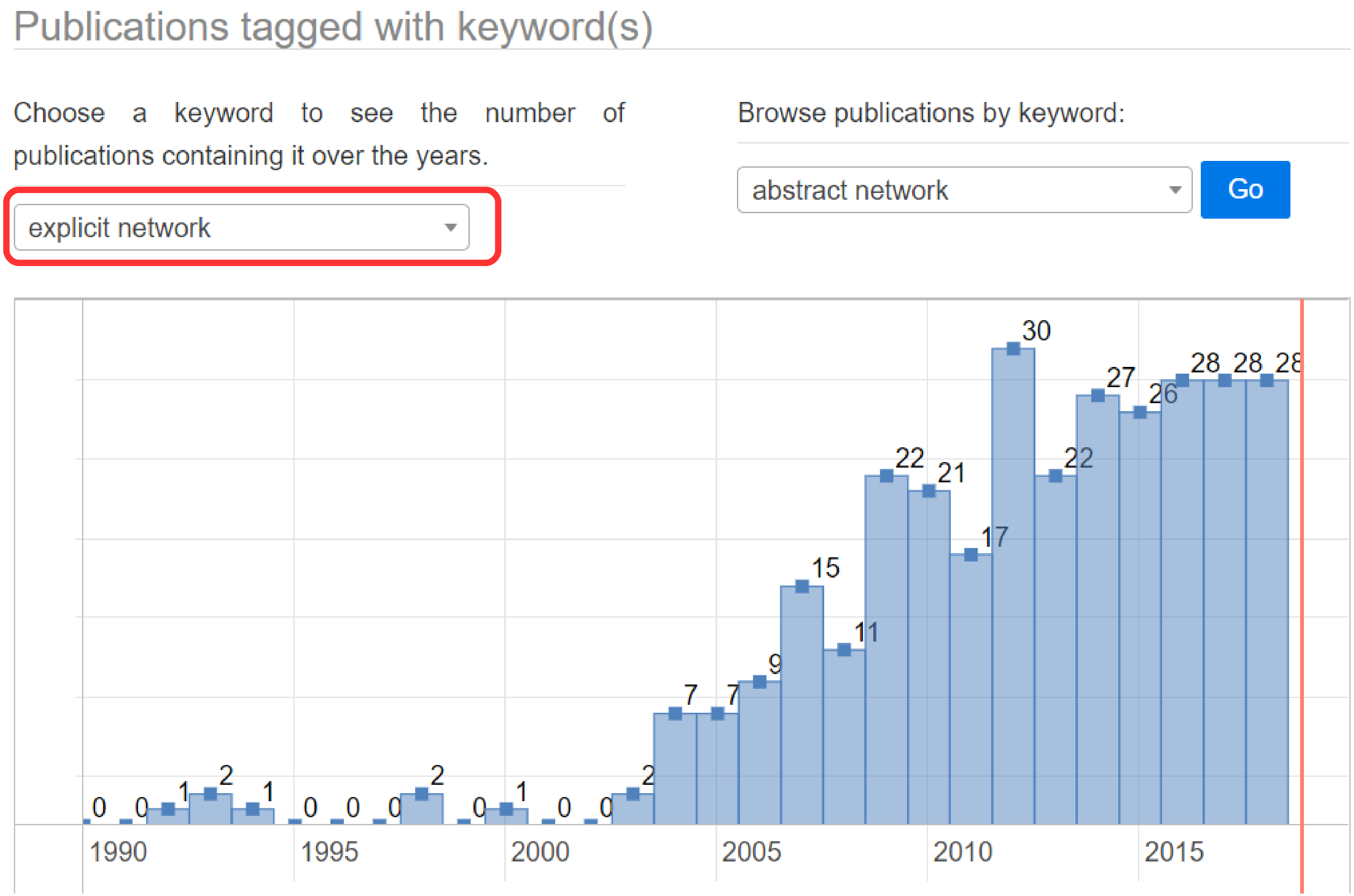
abstract network

Browse publications by keyword:

abstract network

http://phylnet.univ-mlv.fr

Who is who in Phylogenetic Networks?



Publications tagged with keyword(s)

Choose a keyword to see the number of publications containing it over the years.

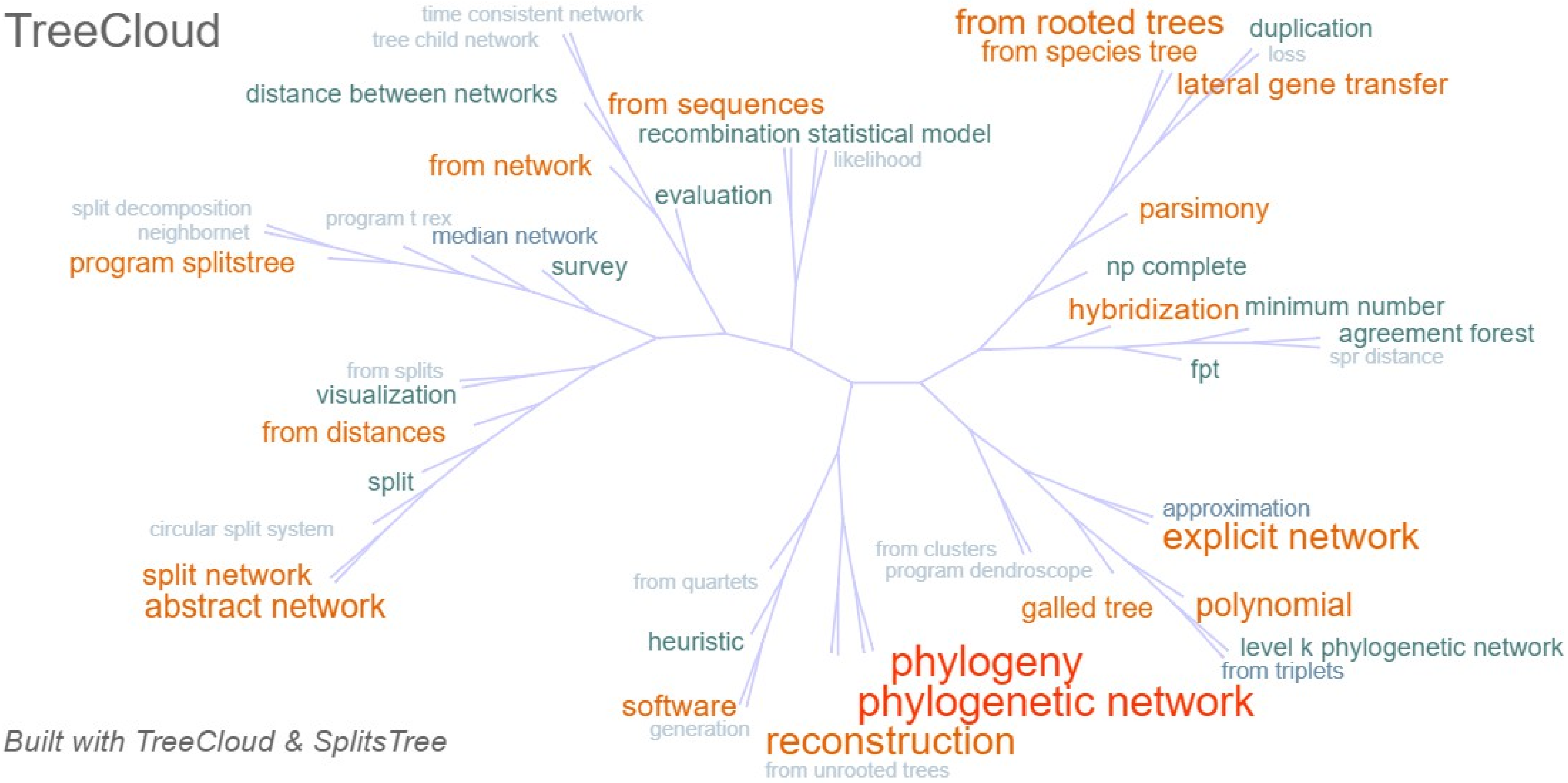
explicit network

Browse publications by keyword:

abstract network

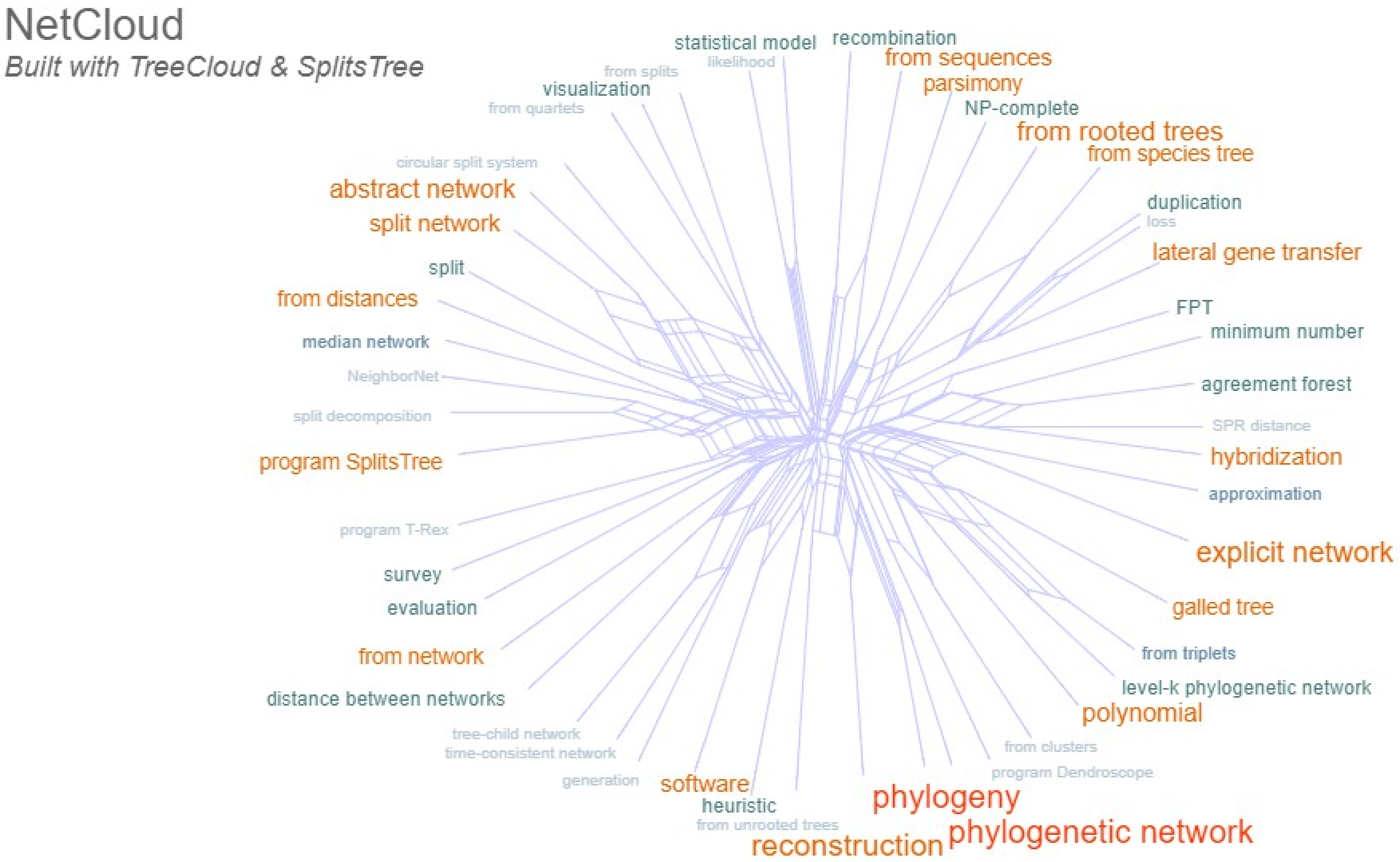
http://phylnet.univ-mlv.fr

Keywords about phylogenetic networks



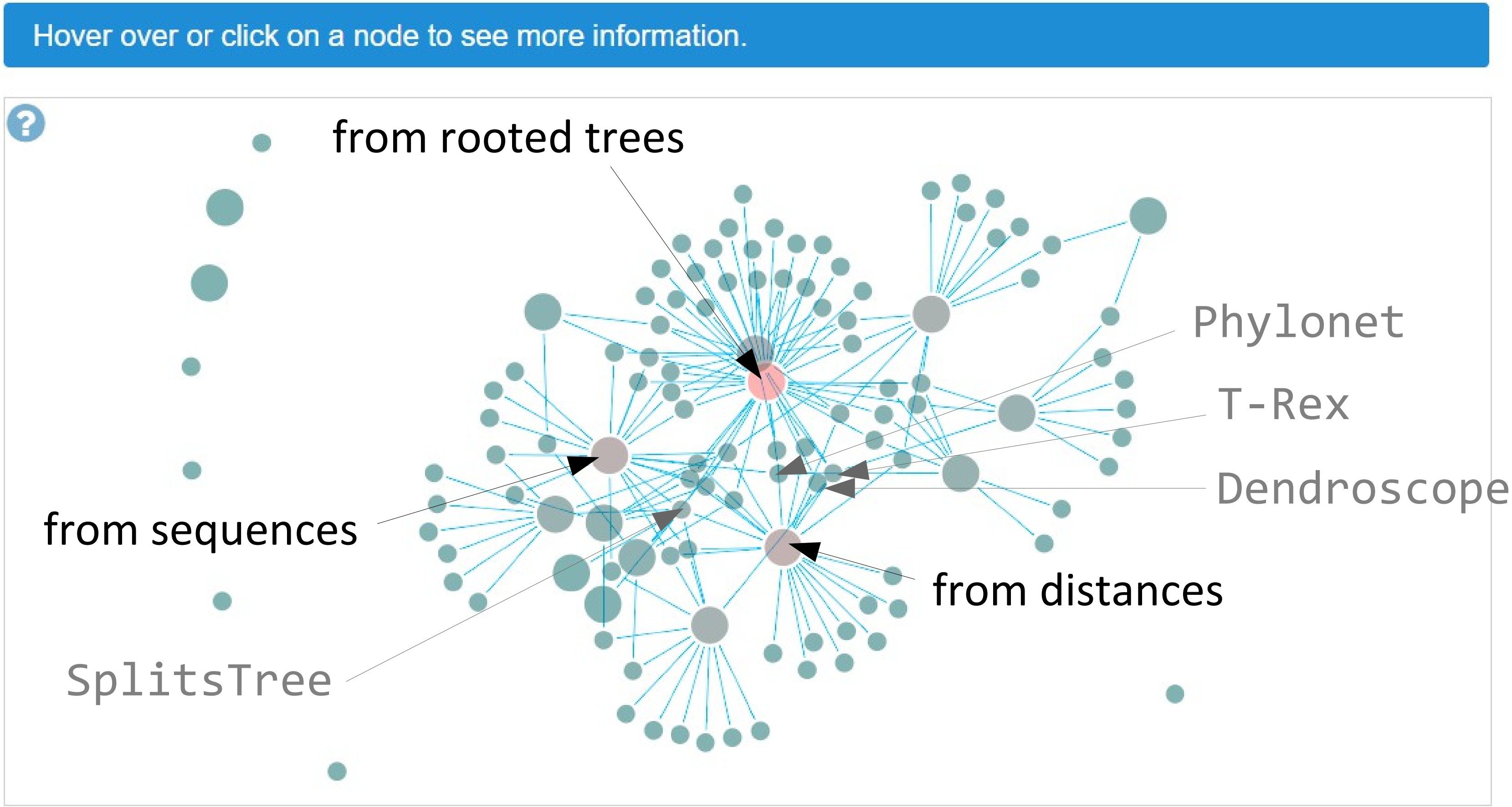
http://phylnet.univ-mlv.fr/keywords.php

Keywords about phylogenetic networks



http://phylnet.univ-mlv.fr/keywords.php

Who is who in Phylogenetic Networks? Software!



Hover over or click on a node to see more information.

Based on BibAdmin by Sergiu Chelcea
+ tag clouds, date histograms, journal lists,
keyword definitions, co-author graphs
(work of Tushar Agarwal)

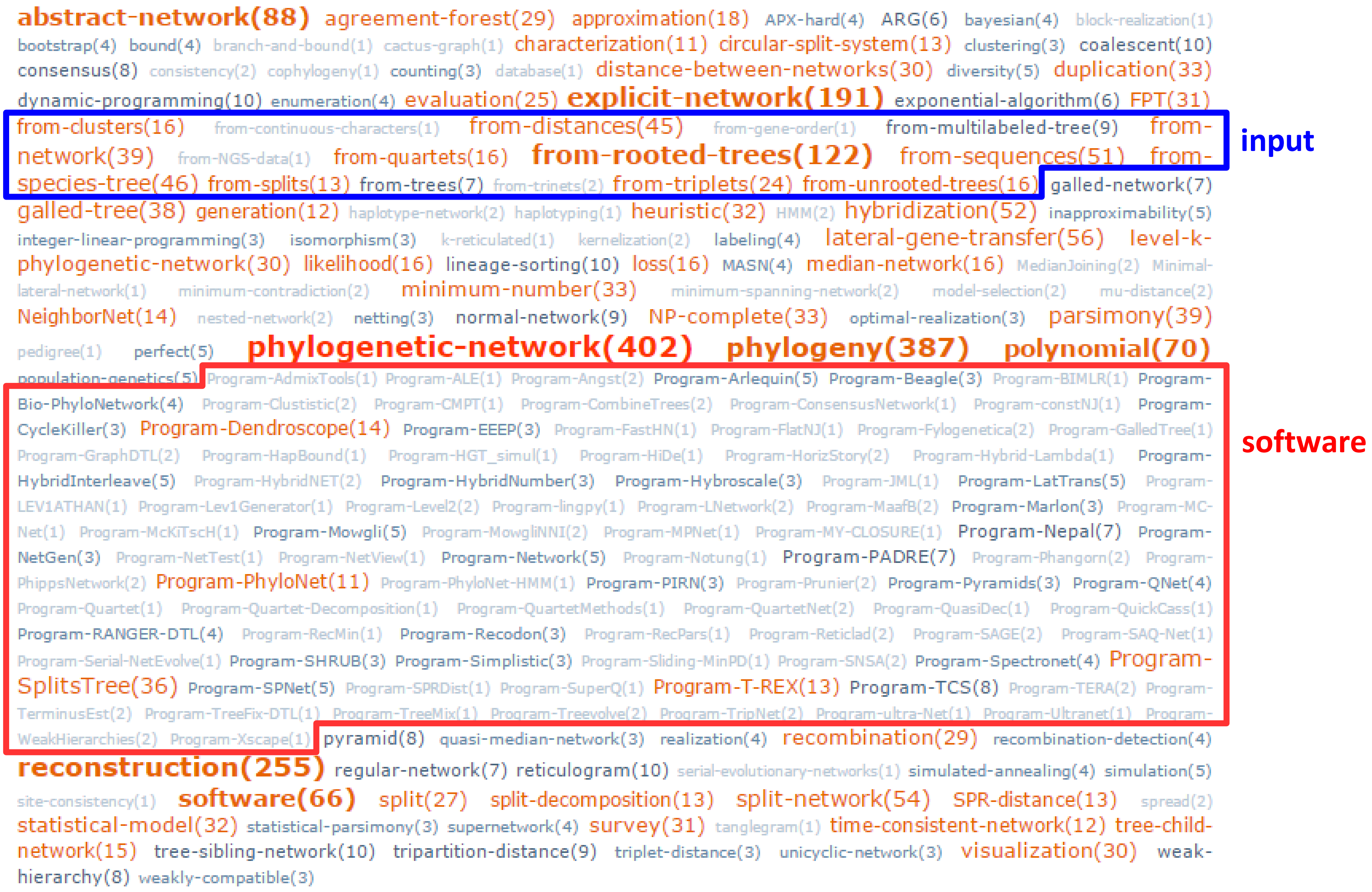
http://phylnet.univ-mlv.fr/show.php?keyword=programs

Who is who in Phylogenetic Networks? Keywords!

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-Lev1Generator(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) supernetwork(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)

http://phylnet.univ-mlv.fr/keywords.php

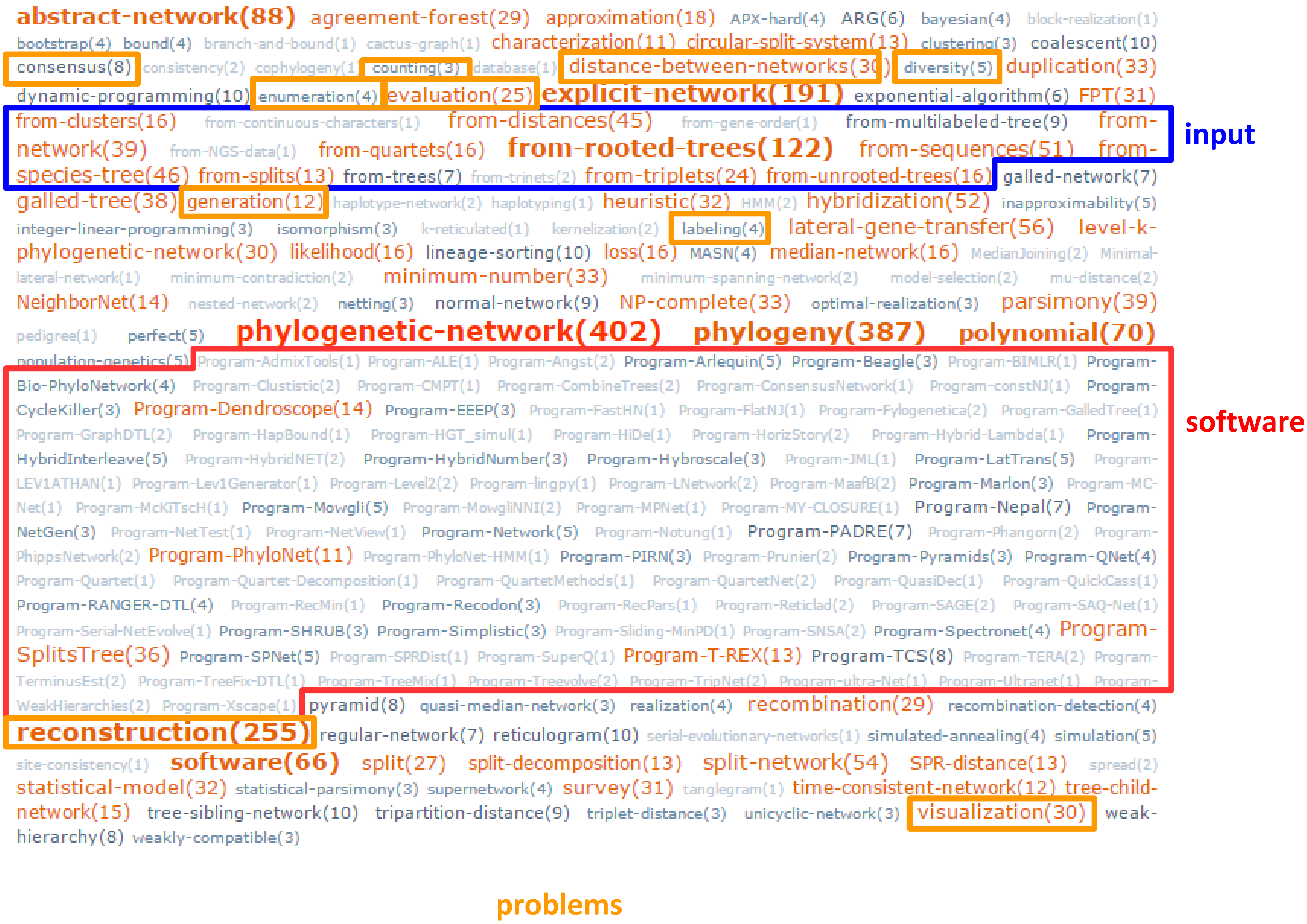
Who is who in Phylogenetic Networks? Keywords!



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-Lev1Generator(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) supernetwork(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)

http://phylnet.univ-mlv.fr/keywords.php

Who is who in Phylogenetic Networks? Keywords!



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-Lev1Generator(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) supernetwork(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? Keywords!

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-Lev1Generator(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) supernetwork(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

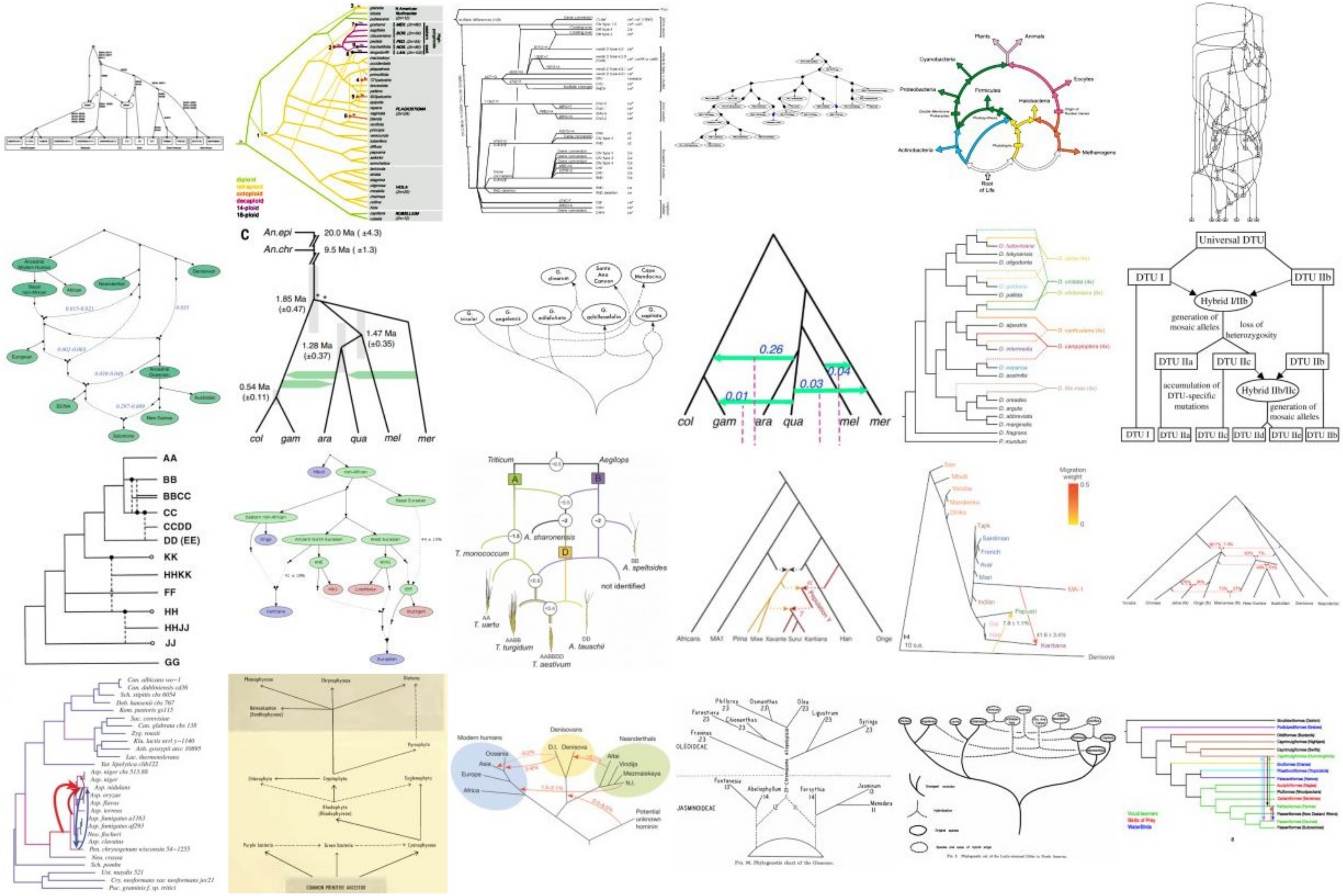
software

problems

algorithmic results

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

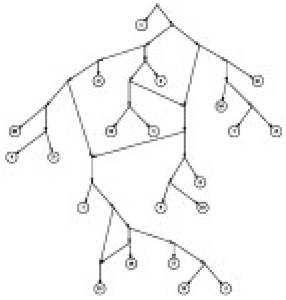
Bandelt & Dress 1992 - Legendre & Makarenkov 2000 - Bryant & Moulton 2002 - Chan, Jansson, Lam & Yiu 2006 - Willems, Tahiri & Makarenkov 2014

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

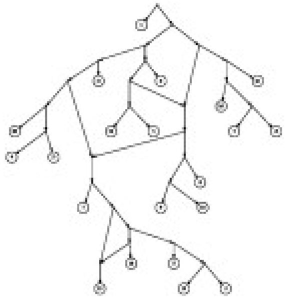
Bandelt & Dress 1992 - Legendre & Makarenkov 2000 - Bryant & Moulton 2002 - Chan, Jansson, Lam & Yiu 2006 - Willems, Tahiri & Makarenkov 2014

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

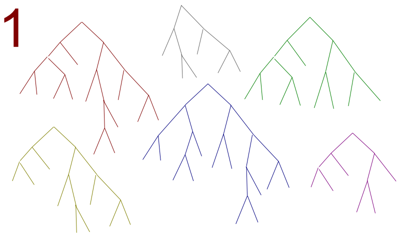
```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2

{gene sequences}

Reconstruction of a tree for each gene present in several species

Guindon & Gascuel, SB, 2003



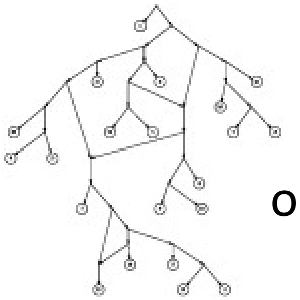
{trees}

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

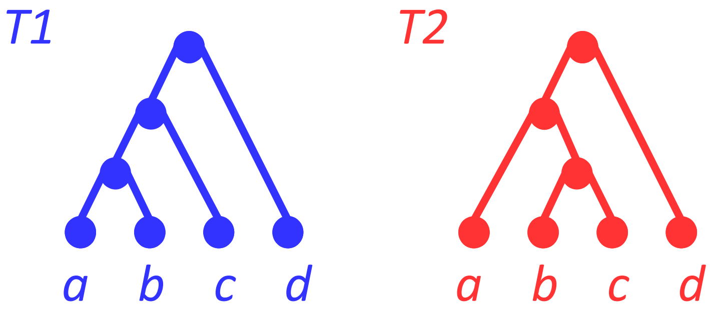
http://doua.prabi.fr/databases/hogenom/home.php?contents=hogenom4

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!

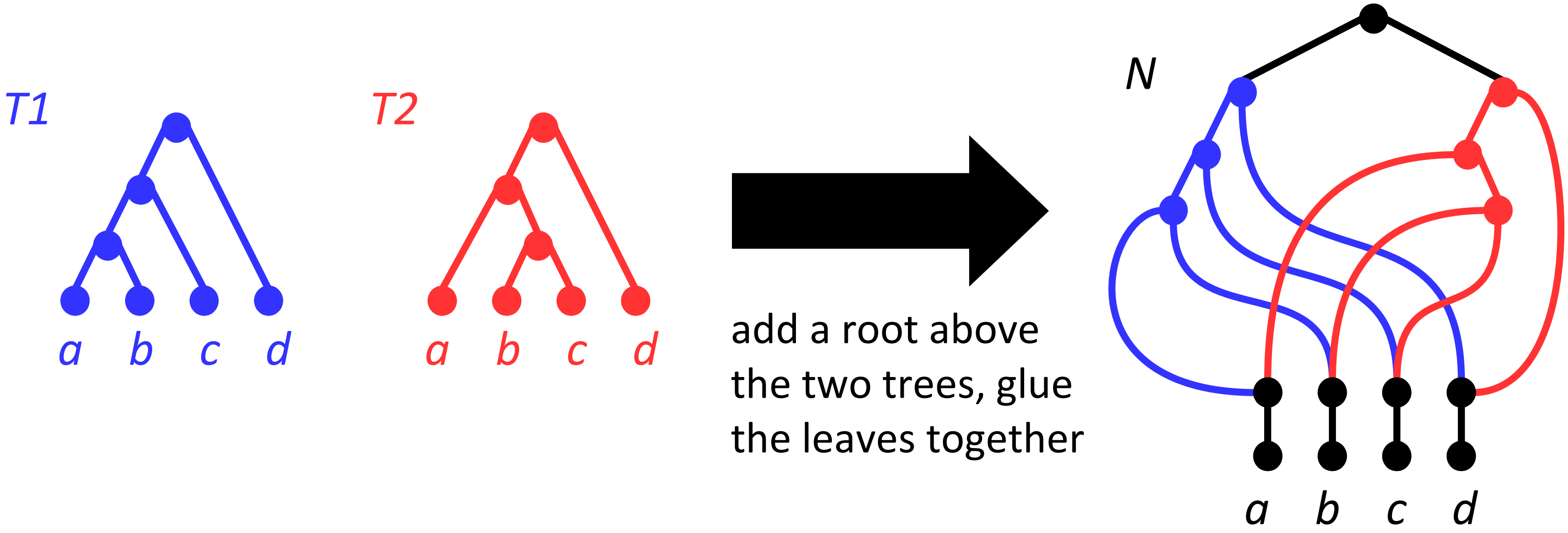


Phylogenetic network reconstruction from trees

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!



add a root above the two trees, glue the leaves together

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

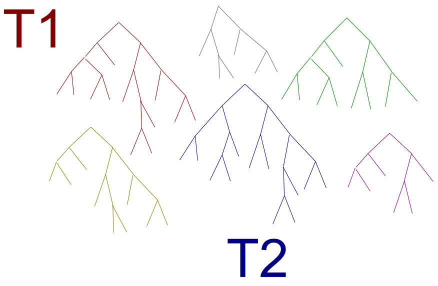
Bordewich & Semple (2007) Discrete Appl Math

Phylogenetic network reconstruction from trees

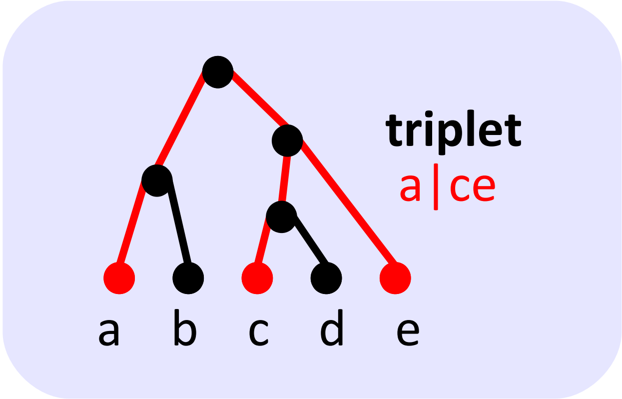
```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

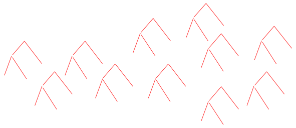
G1 G2

{gene sequences}



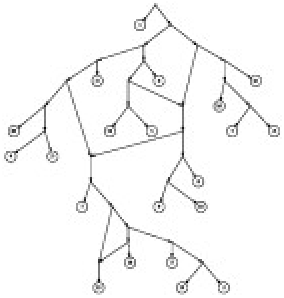
{trees}





{triplets}

explicit network

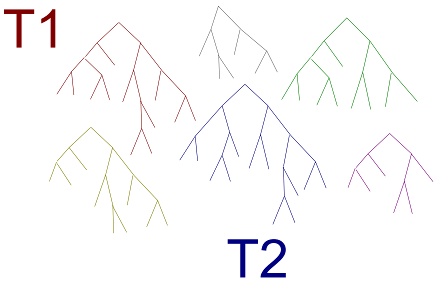


optimal super-network *N*:
- contains the input triplets
- has the smallest number of reticulations

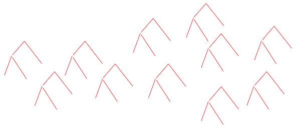
Phylogenetic network reconstruction from trees

	G1	G2
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*

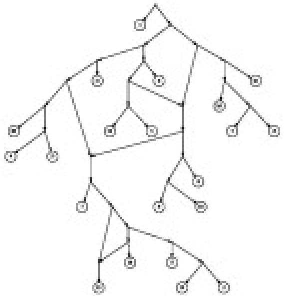
{gene sequences}



{trees}



{triplets}



explicit network

optimal super-network *N*:

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

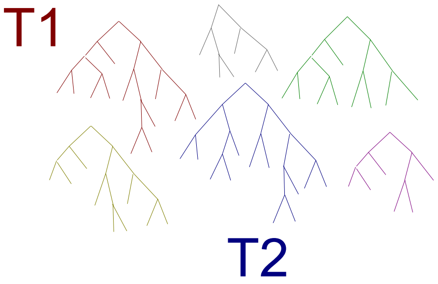
→ still NP-hard

Phylogenetic network reconstruction from trees

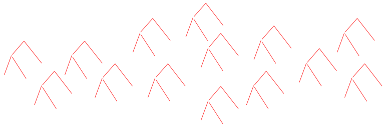
```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2

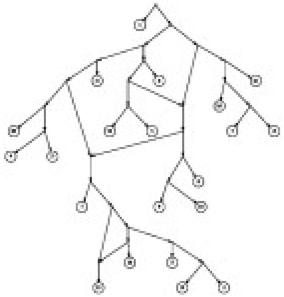
{gene sequences}



{trees}



{triplets} ("dense" set)



explicit network

super-network *N*:

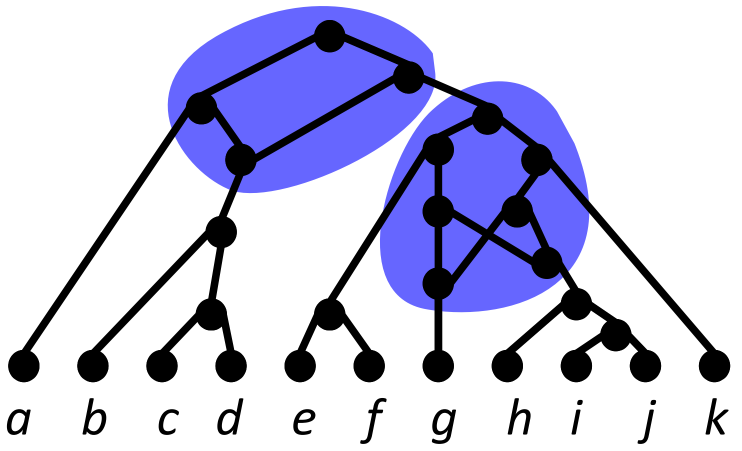
- contains the input triplets
- has level at most *k* (i.e. close to a tree)

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

Outline

- An introduction to phylogenetic networks
- Classes of phylogenetic networks
- Counting problems on phylogenetic networks

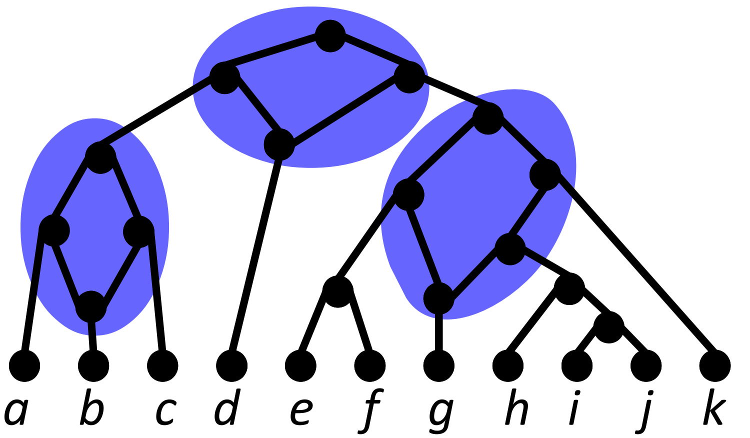
Classes of phylogenetic networks: level-k networks



level-2 network

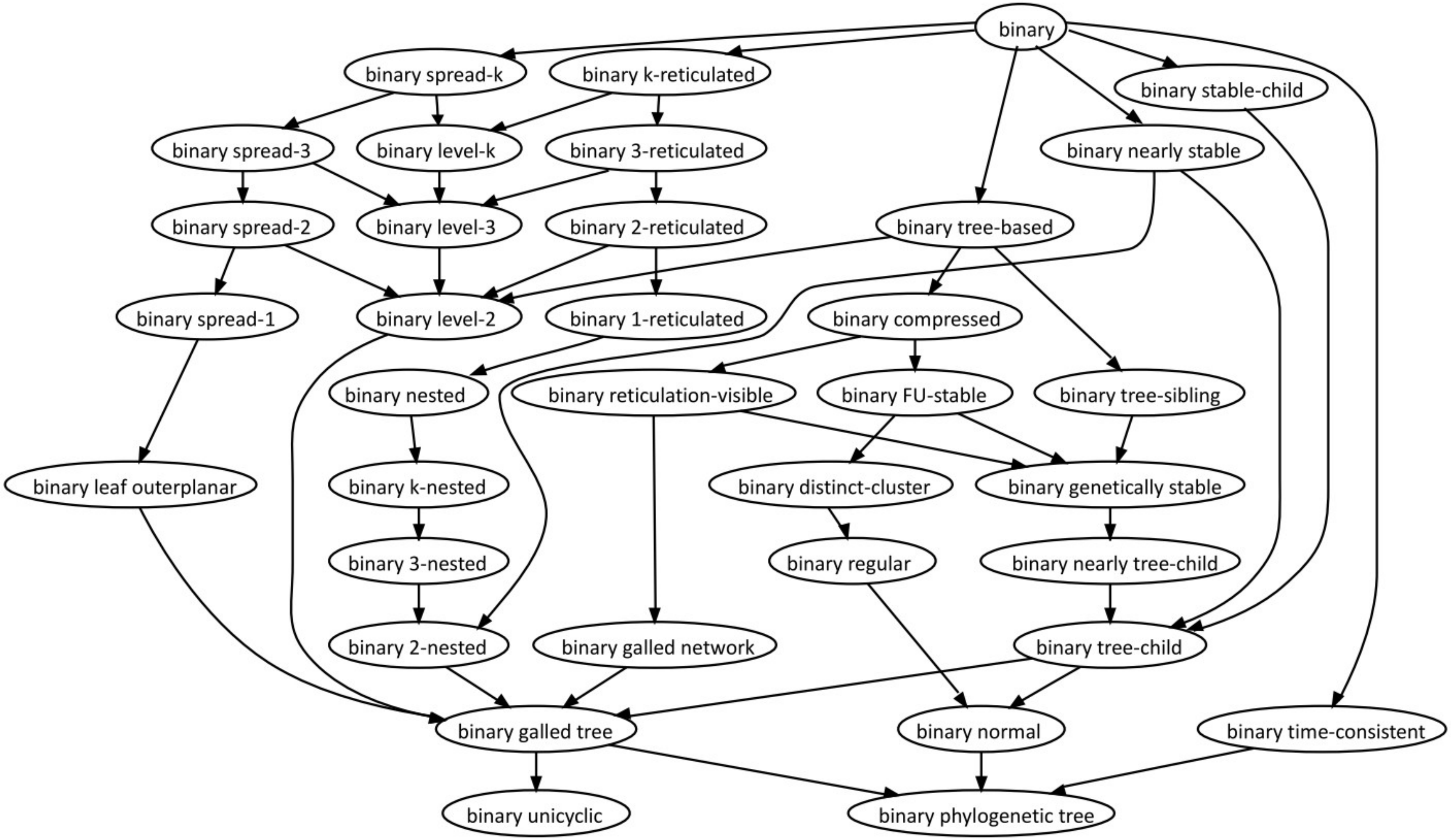
level =
maximum number of reticulations per **blob** (=bridgeless component of the underlying undirected graph).

level-1 network
(“*galled tree*”)



Classes of phylogenetic networks: ISIPhyNC

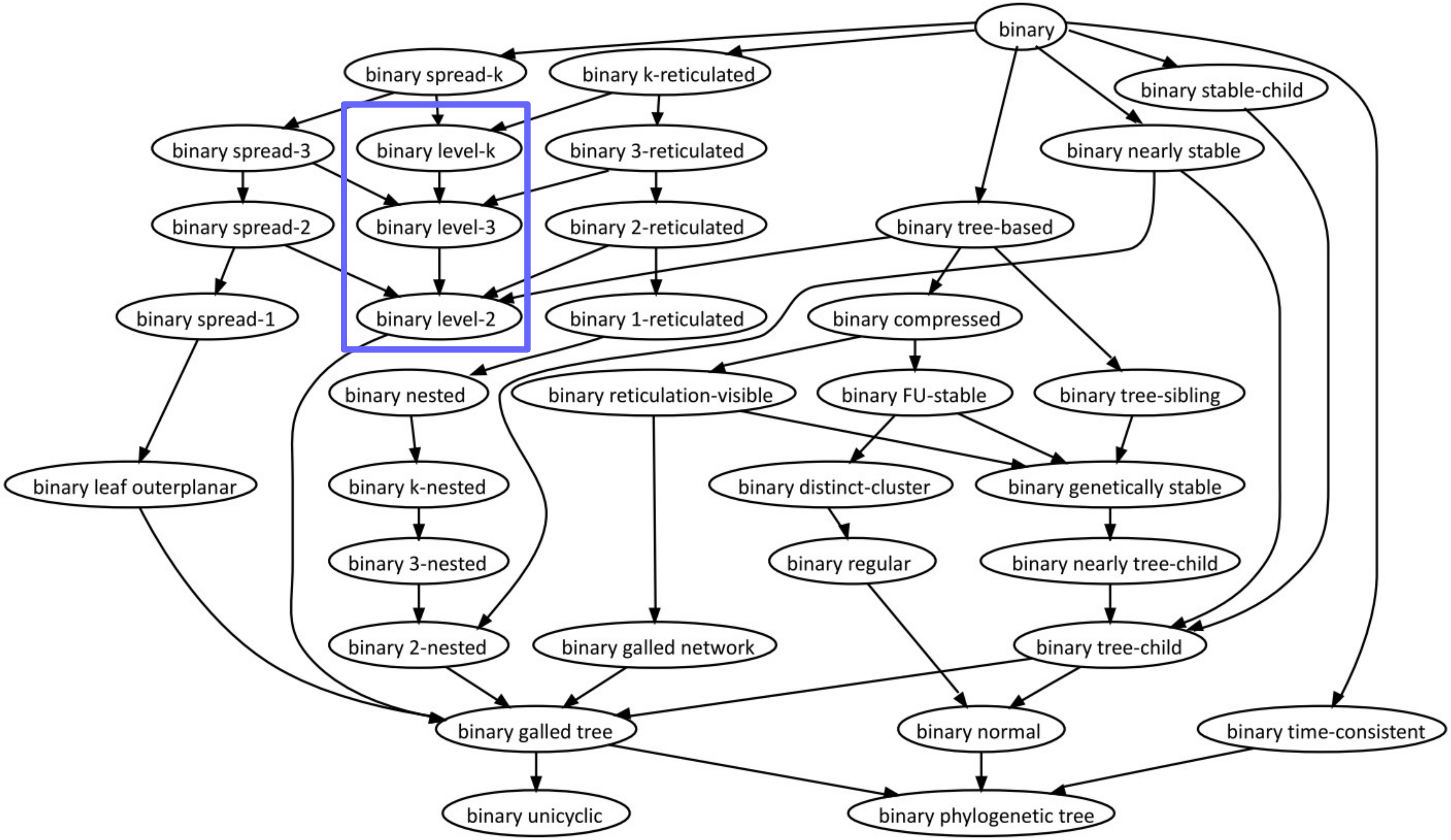
Information System on Inclusions of Phylogenetic Network Classes



joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/

Classes of phylogenetic networks: ISIPhyNC

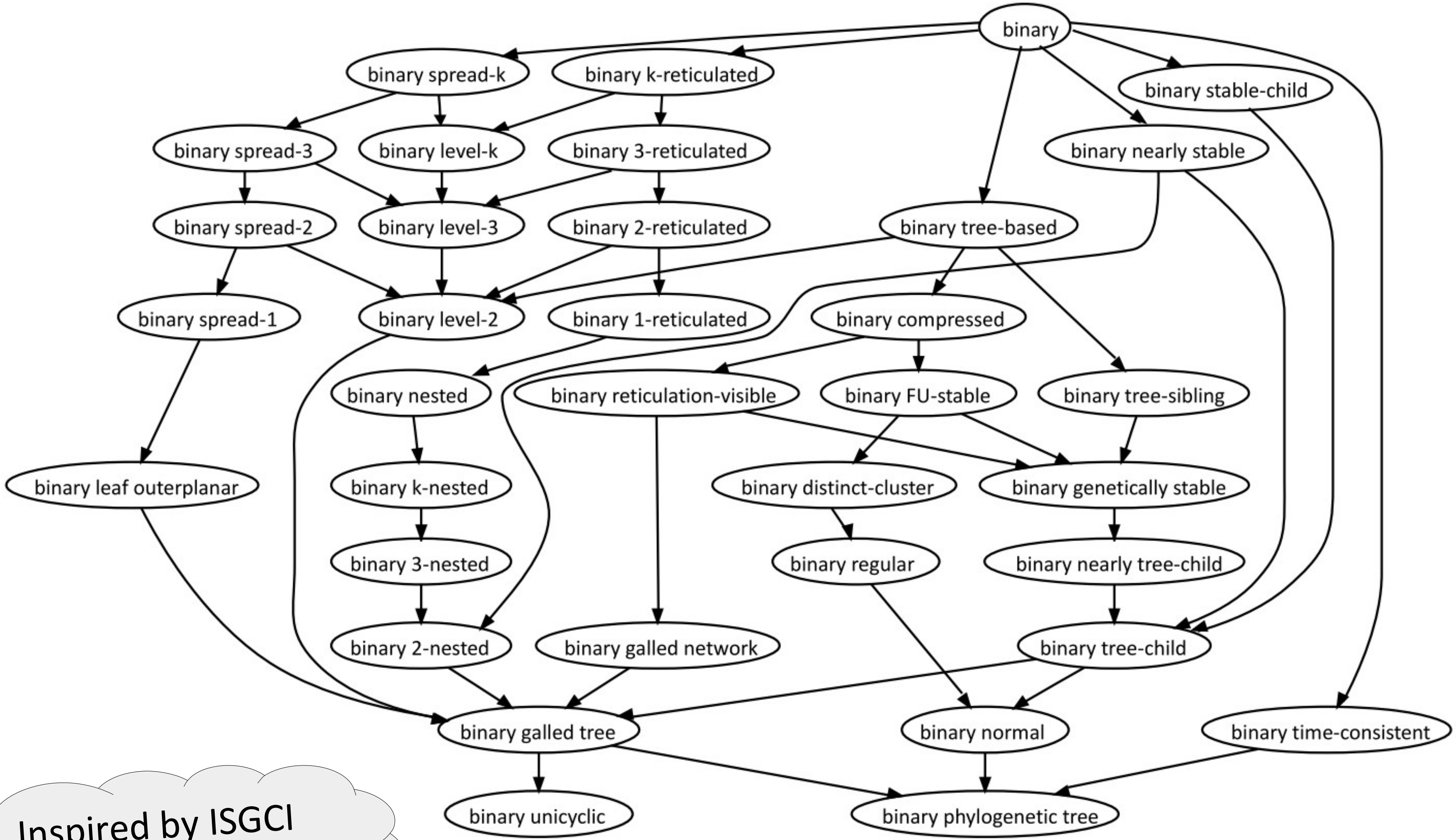
Information System on Inclusions of Phylogenetic Network Classes



joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/

Classes of phylogenetic networks: ISIPhyNC

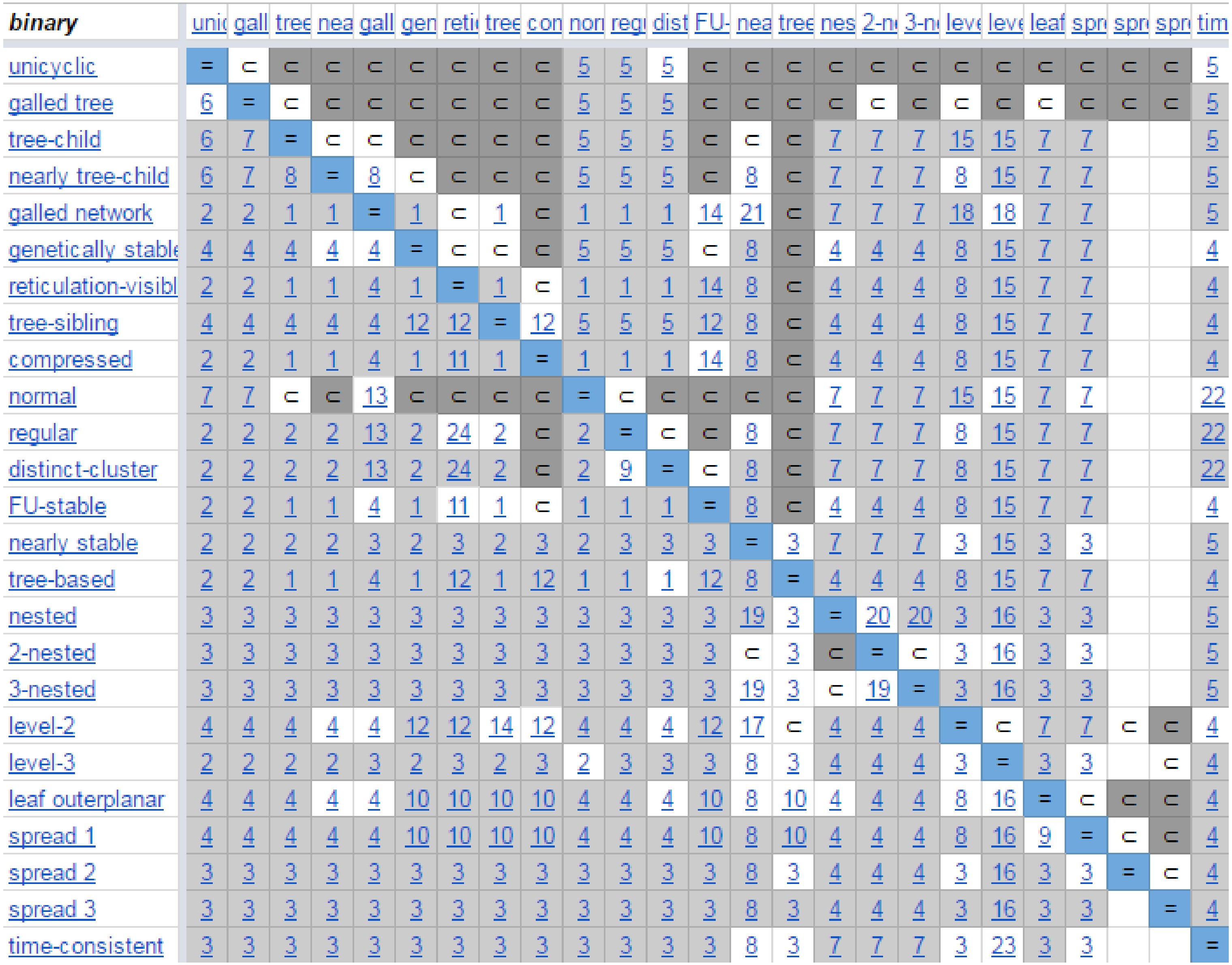
Information System on Inclusions of Phylogenetic Network Classes



Inspired by ISGCI
graphclasses.org

joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/

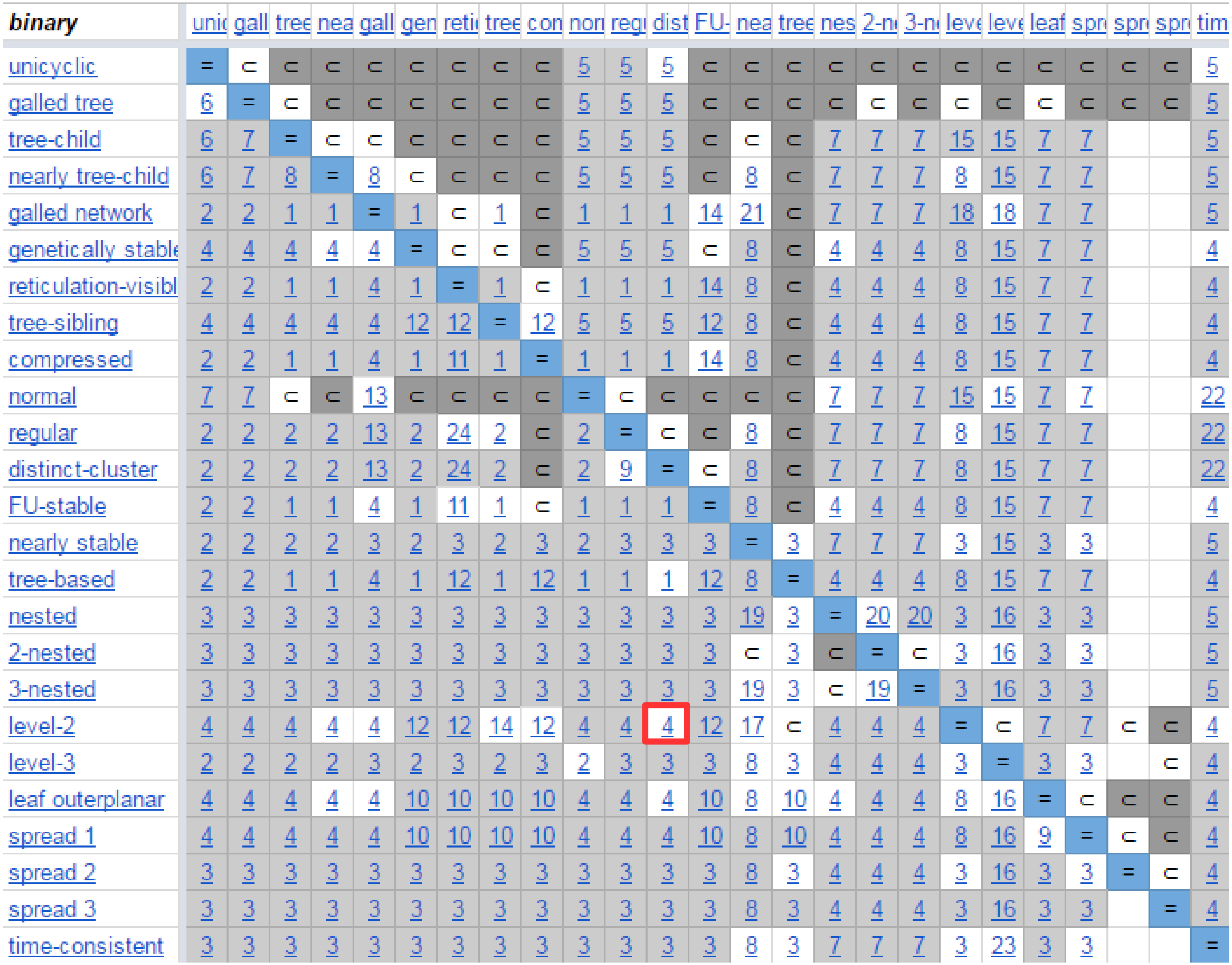
Classes of phylogenetic networks: inclusions



binary	unic	gall	tree	nea	gall	gen	reti	tree	con	nor	regi	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

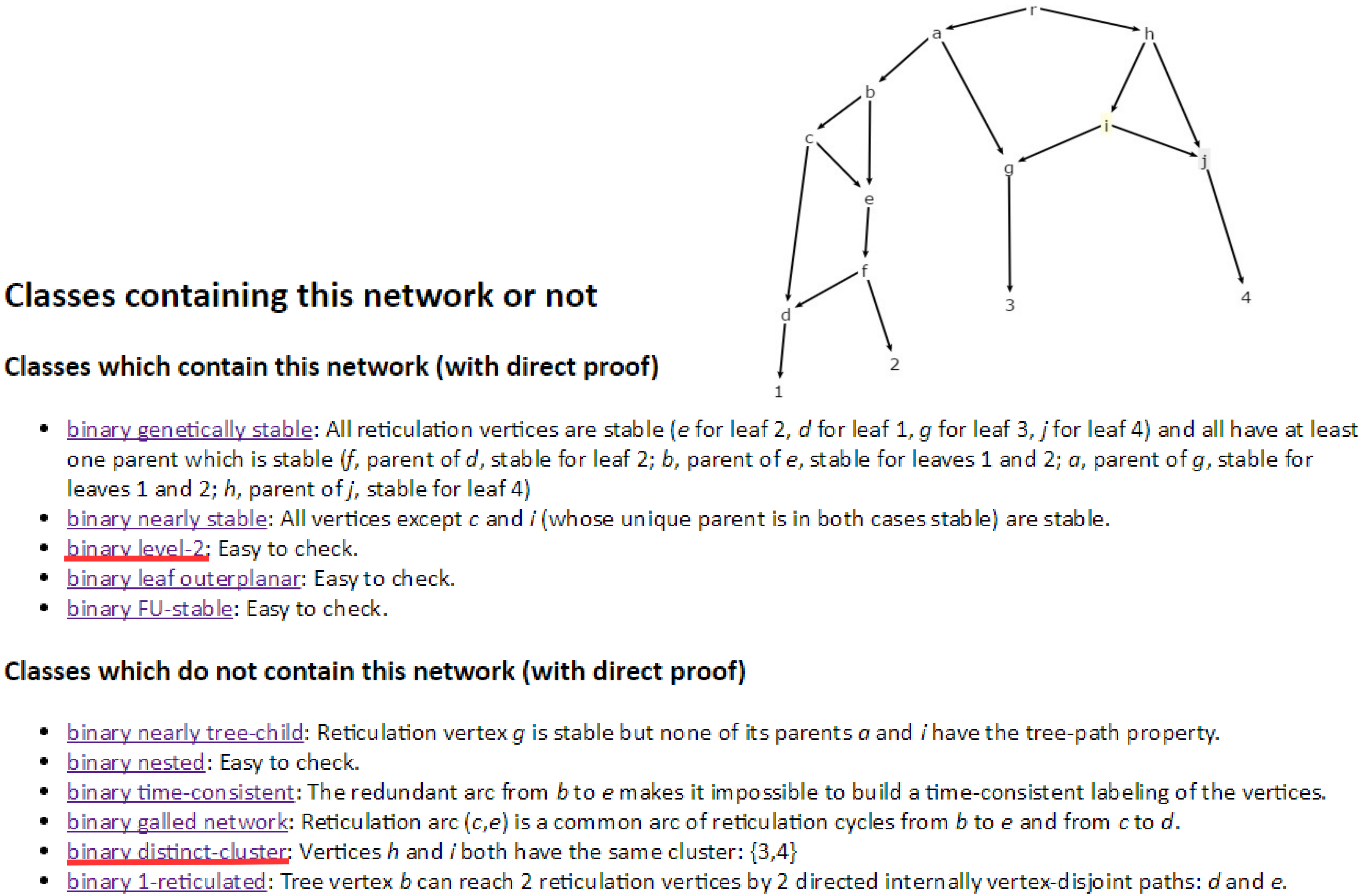
Classes of phylogenetic networks: inclusions



binary	uni	gall	tree	nea	gall	gen	reti	tree	con	nor	reg	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.

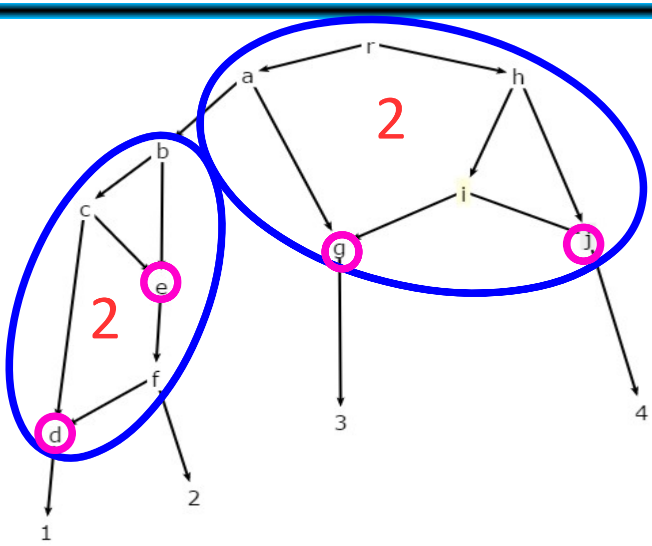
Classes which do not contain this network (with direct proof)

- 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

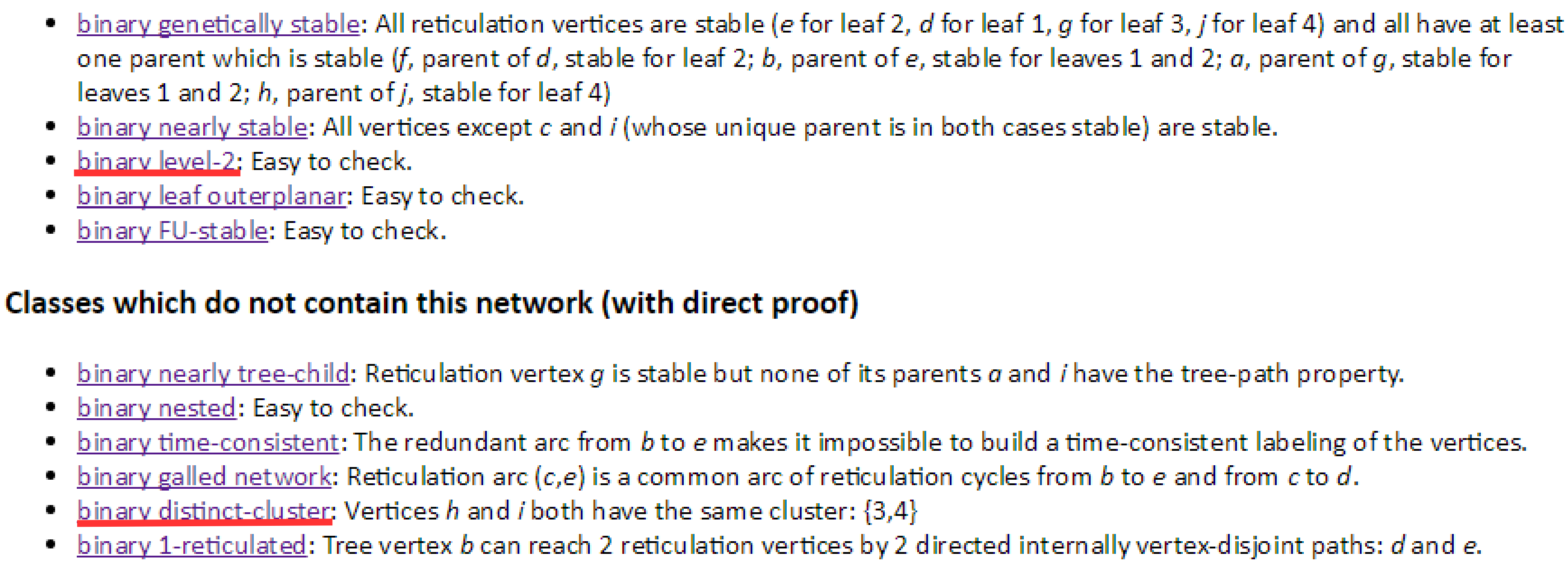
Classes of phylogenetic networks: inclusions

level = maximum number of **reticulation vertices** among all **bridgeless components** in the network



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.

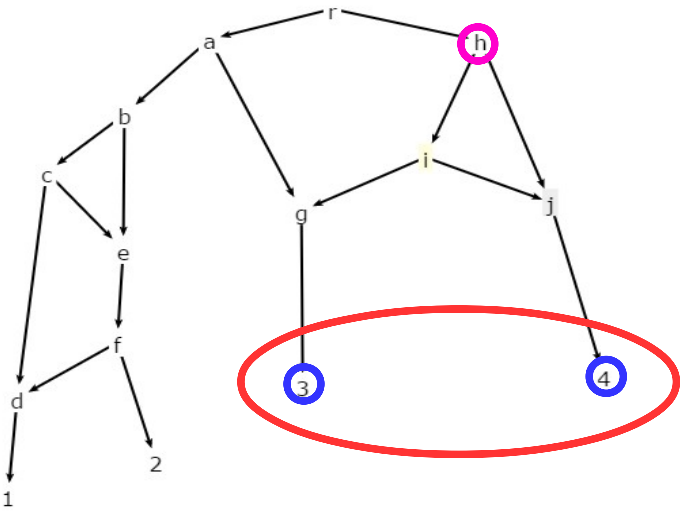
Classes which do not contain this network (with direct proof)

- 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*.

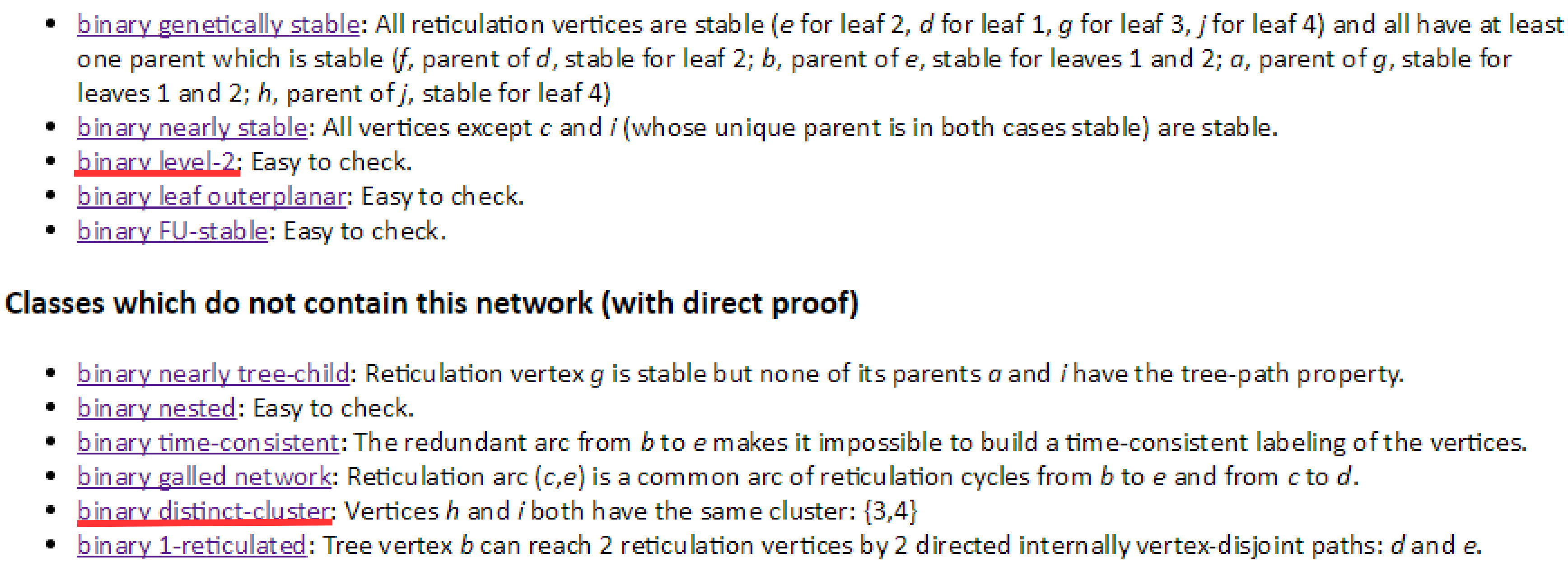
joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/network.php?id=4

Classes of phylogenetic networks: inclusions

cluster = set of **leaves** below a **vertex**
distinct-cluster = all vertices (except reticulation vertices with their child) have distinct clusters



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.

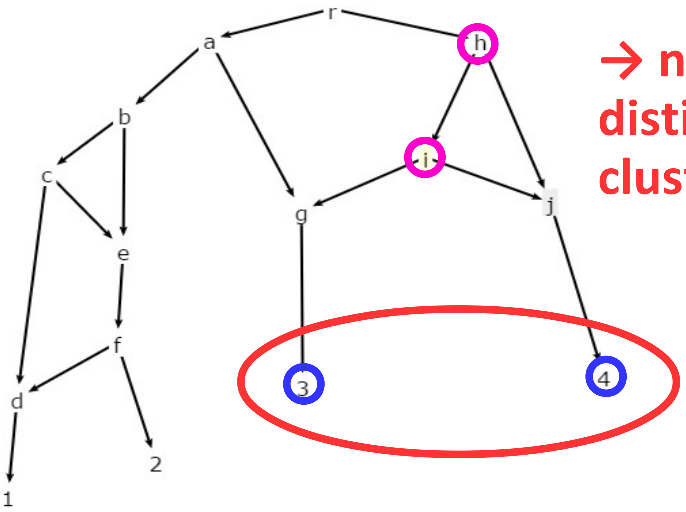
Classes which do not contain this network (with direct proof)

- 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

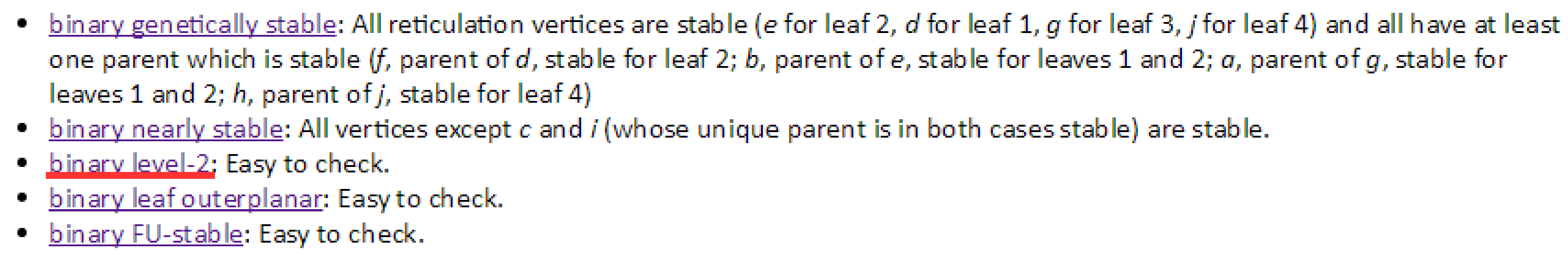
cluster = set of **leaves** below a **vertex**
distinct-cluster = all vertices (except reticulation vertices with their child) have distinct clusters



→ **not distinct-cluster**

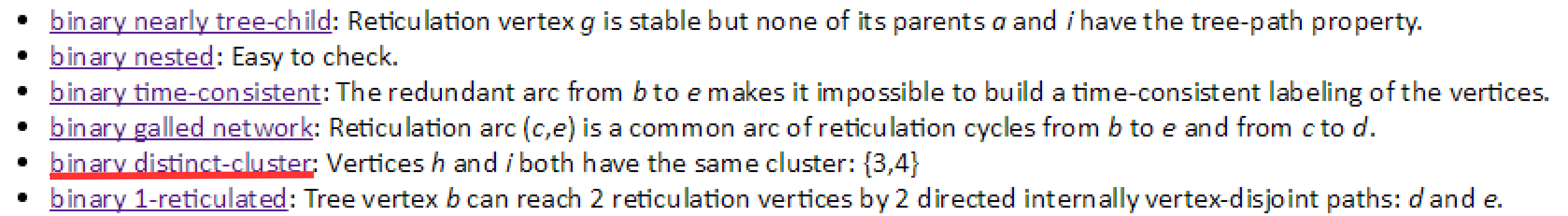
Classes containing this network or not

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.

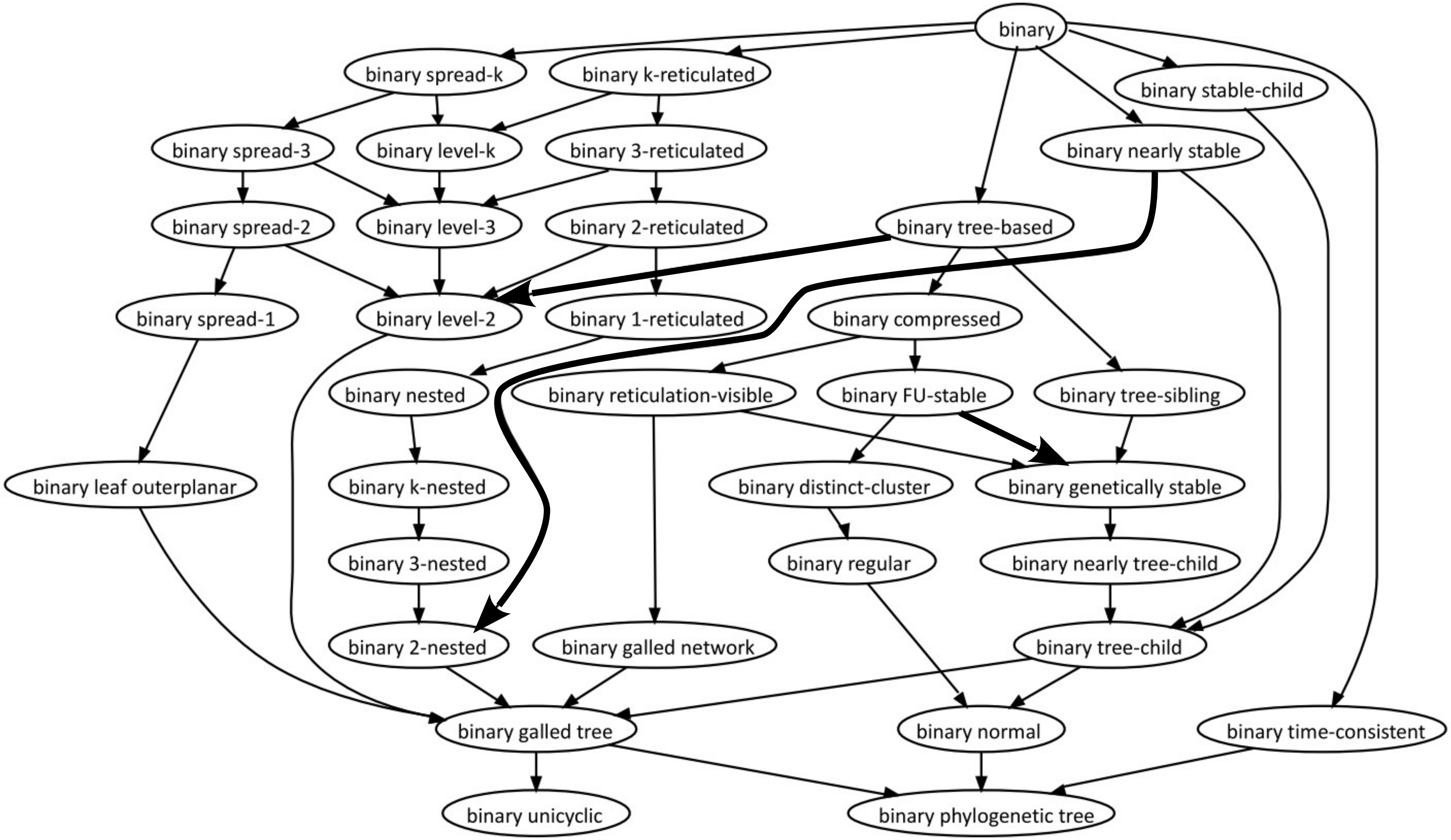
Classes which do not contain this network (with direct proof)



- 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

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-Lev1Generator(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) supernetwork(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

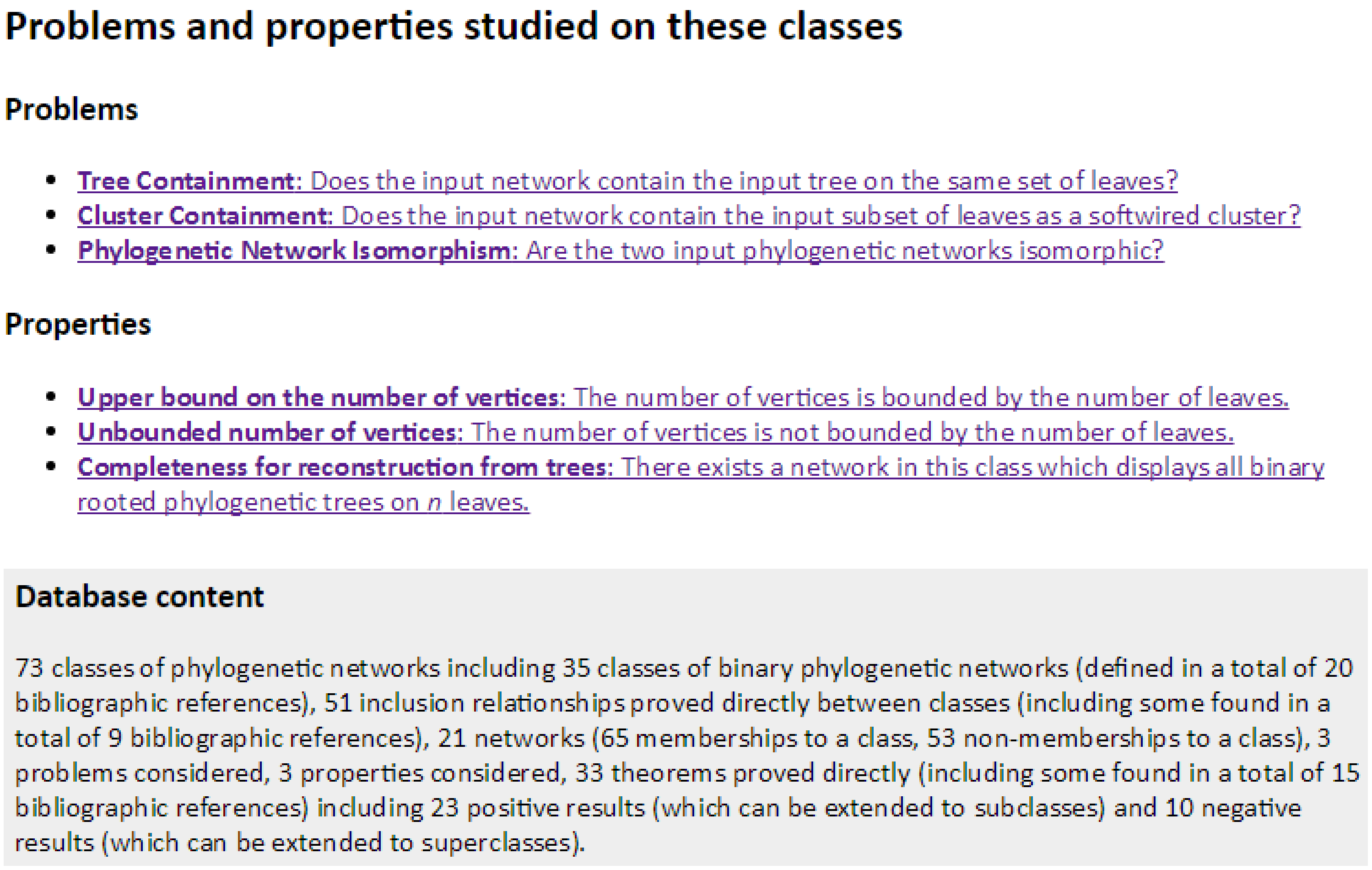
software

classes

problems

algorithmic properties

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

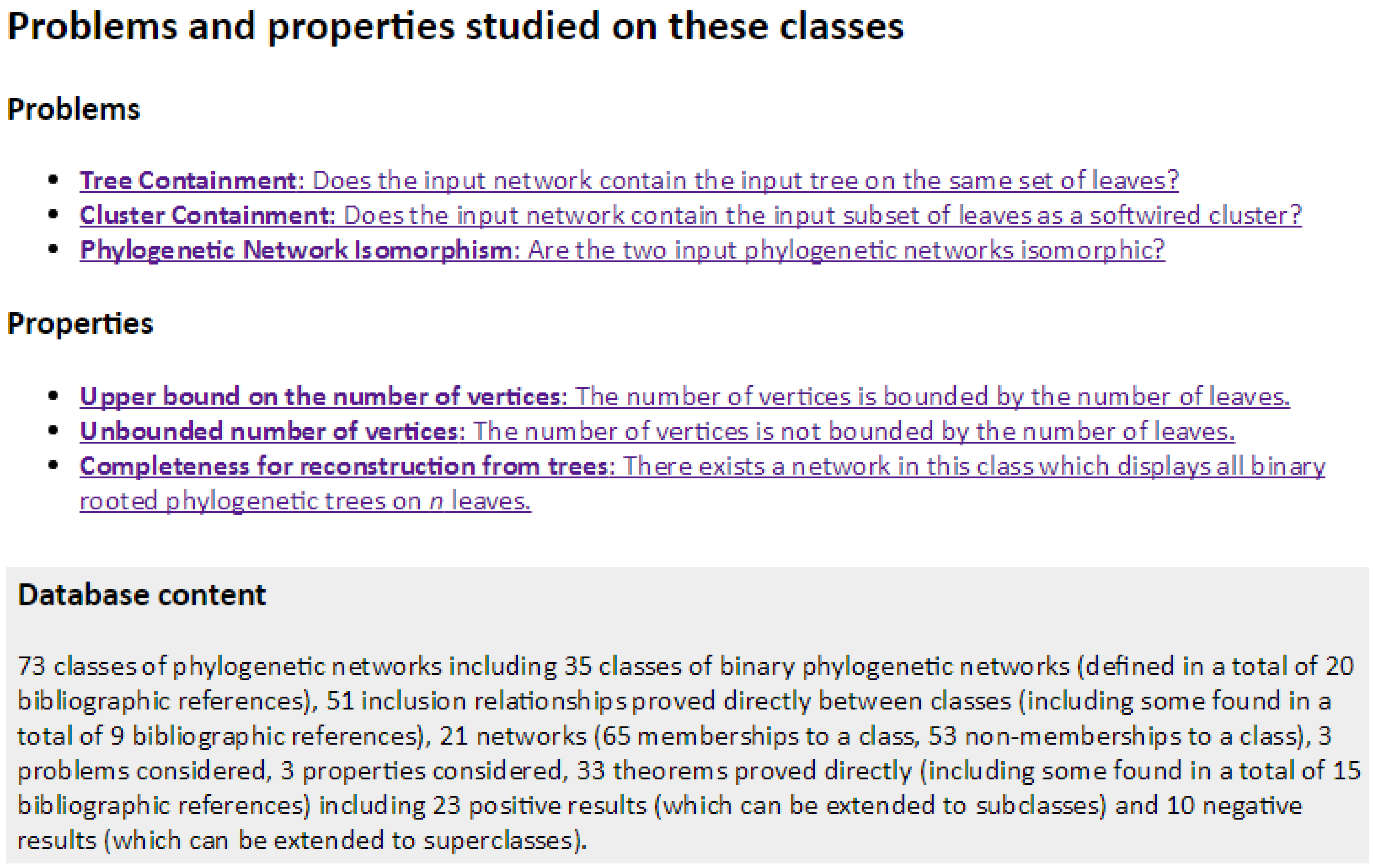
- **Upper bound on the number of vertices:** The number of vertices is bounded by the number of leaves.
- **Unbounded number of vertices:** The number of vertices is not bounded by the number of leaves.
- **Completeness for reconstruction from trees:** There exists a network in this class which displays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

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*?

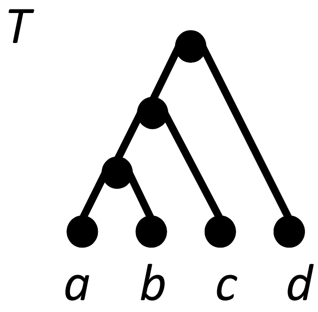
The Tree Containment Problem (T.C.P.)

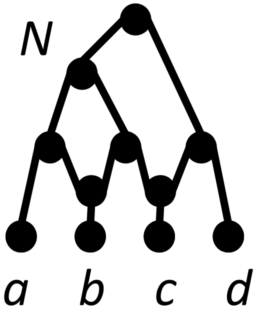
Input: A binary phylogenetic network *N* and

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*?





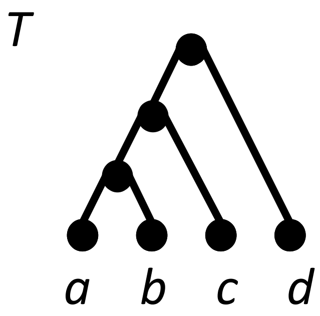
The Tree Containment Problem (T.C.P.)

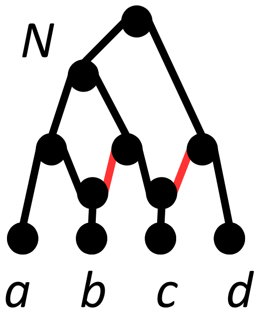
Input: A binary phylogenetic network *N* and

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*?

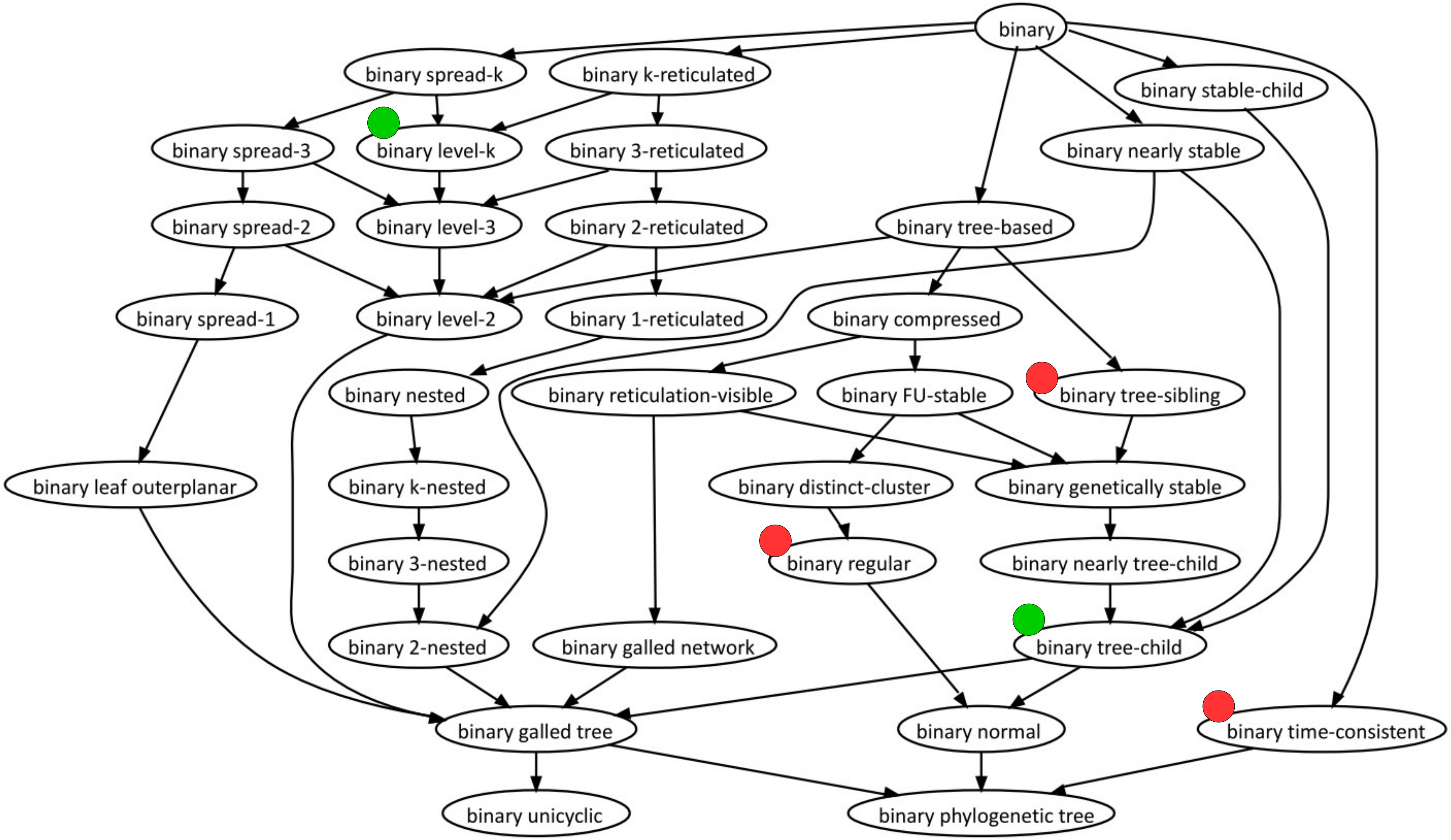
- **NP-complete** in general (Kanj, Nakhleh, Than & Xia, 2008)

- **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)

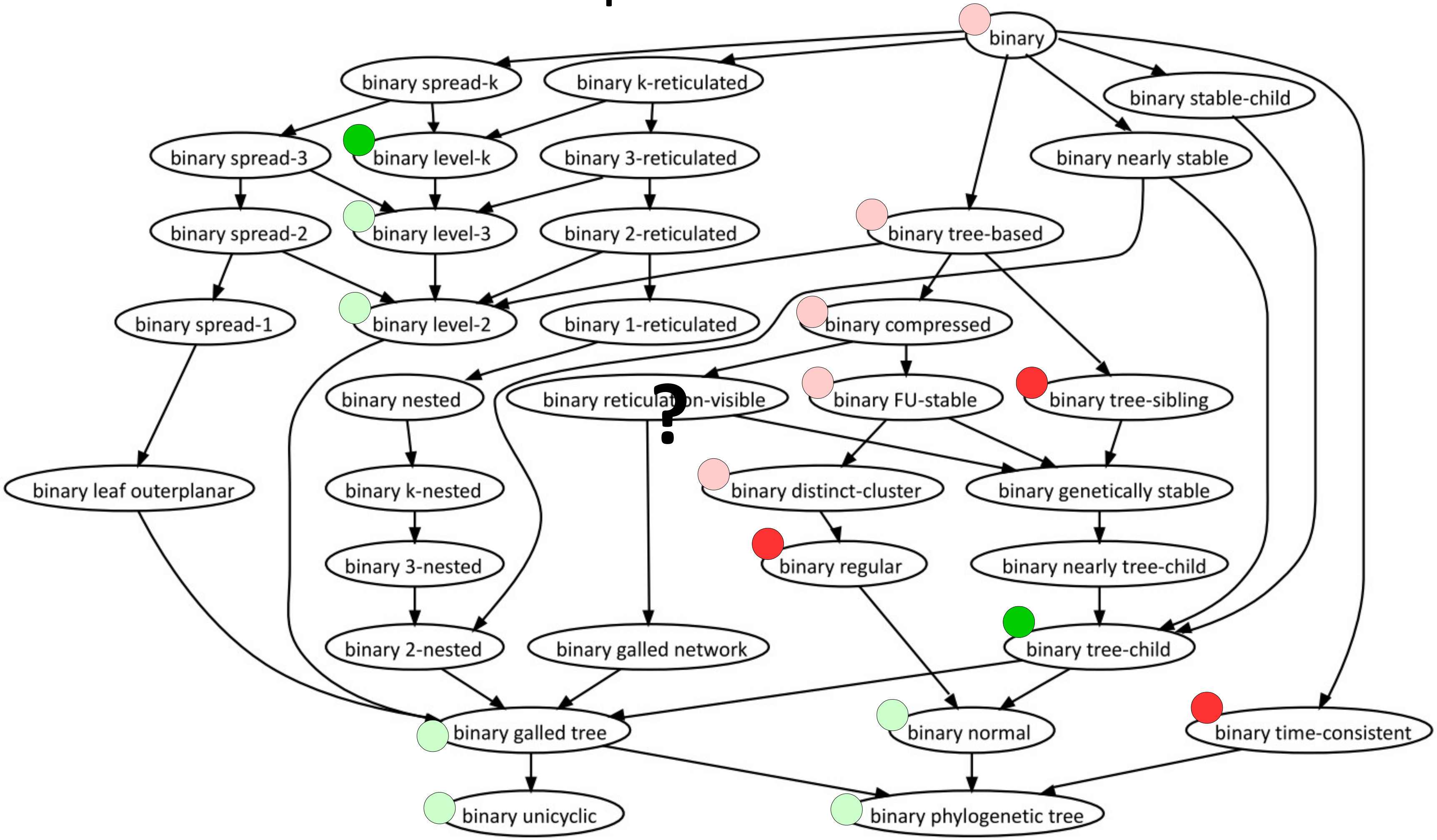
Classes of phylogenetic networks and the T.C.P.

State of the art in 2011: available proofs

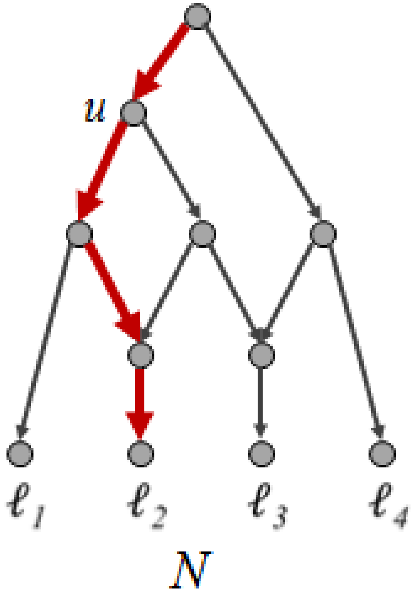


Classes of phylogenetic networks and the T.C.P.

State of the art in 2011: available proofs + results deduced from inclusions

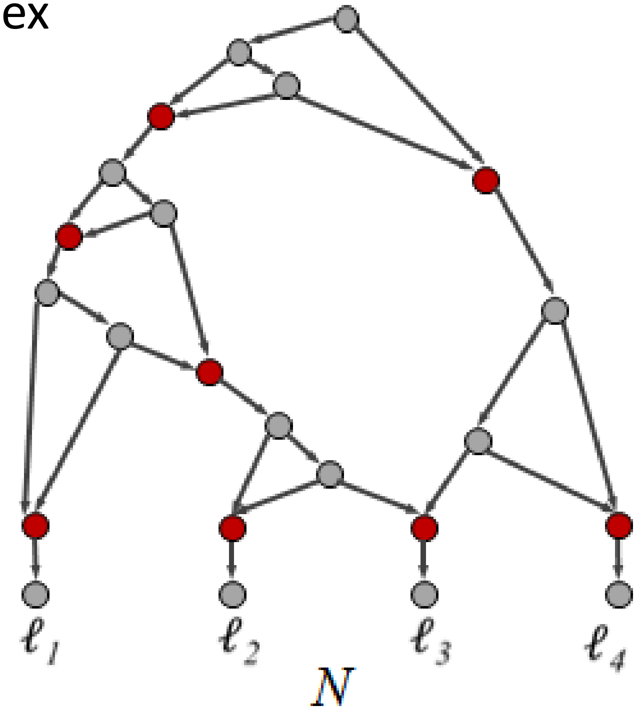


Reticulation-visible and nearly-stable networks



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

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 reticulation-visible then: #{reticulation vertices of *N*} $\leq 4(n-1)$
#{vertices of *N*} $\leq 9n$

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

Theorem 3: Considering a longest path in *N* (nearly stable), it is possible, in constant time:

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

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

Number of reticulations of a reticulation-visible network

Decompose *N* into **2*n*-2 paths**:

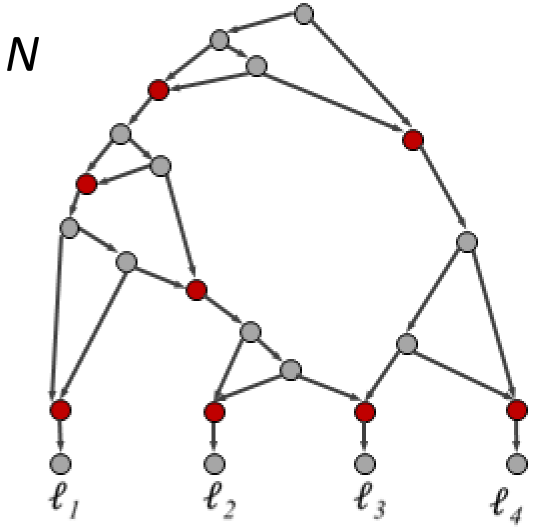
- remove one reticulation arc per reticulation,

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 2*n*-2 arcs

We can prove (technical) that:

each path contains at most 2 reticulation vertices



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

Number of reticulations of a reticulation-visible network

Decompose *N* into **2*n*-2 paths**:

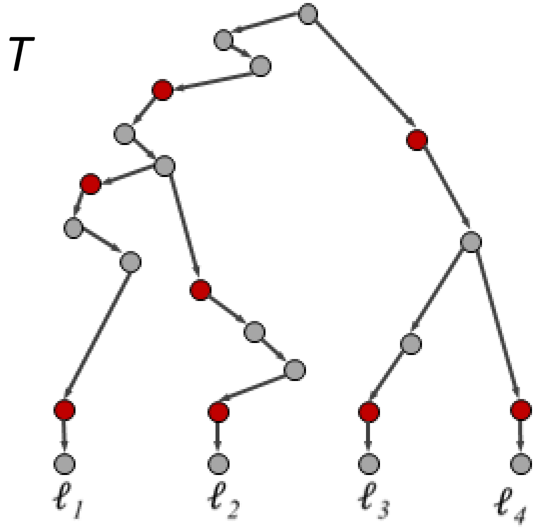
- remove one reticulation arc per reticulation,

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 2*n*-2 arcs

We can prove (technical) that:

each path contains at most 2 reticulation vertices



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

Number of reticulations of a reticulation-visible network

Decompose *N* into **2*n*-2 paths**:

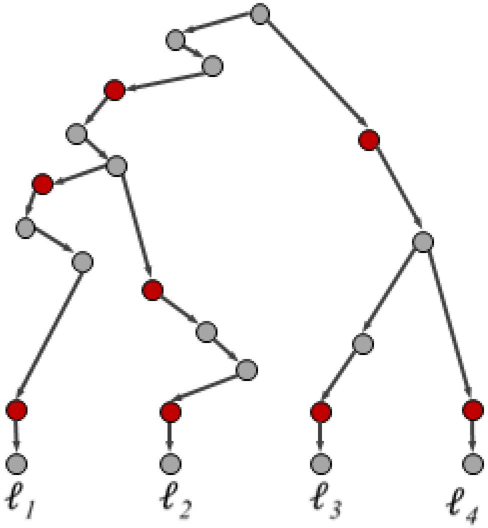
- remove one reticulation arc per reticulation,
ensuring we get no « dummy leaf », to get a tree *T* with *n* leaves

- summarize *T* into a rooted binary tree *T'* with *n* leaves... and 2*n*-2 arcs

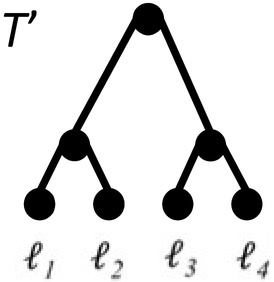
We can prove (technical) that:

each path contains at most 2 reticulation vertices

T



T'



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

Number of reticulations of a reticulation-visible network

Decompose *N* into **2*n*-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 2*n*-2 arcs

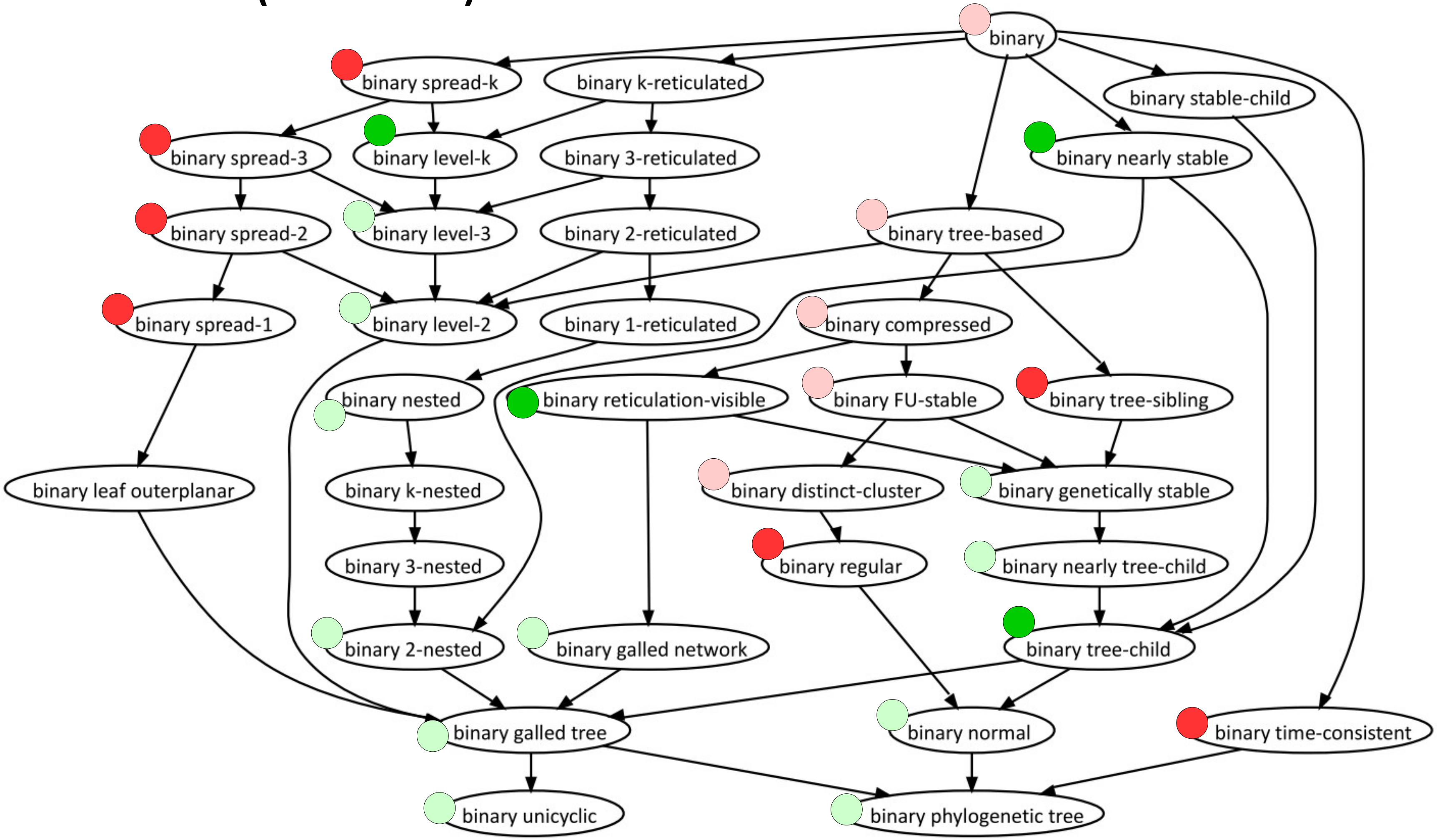
We can prove (technical) that:
each path contains at most 2 reticulation vertices

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

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

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

Recent results (since 2015):



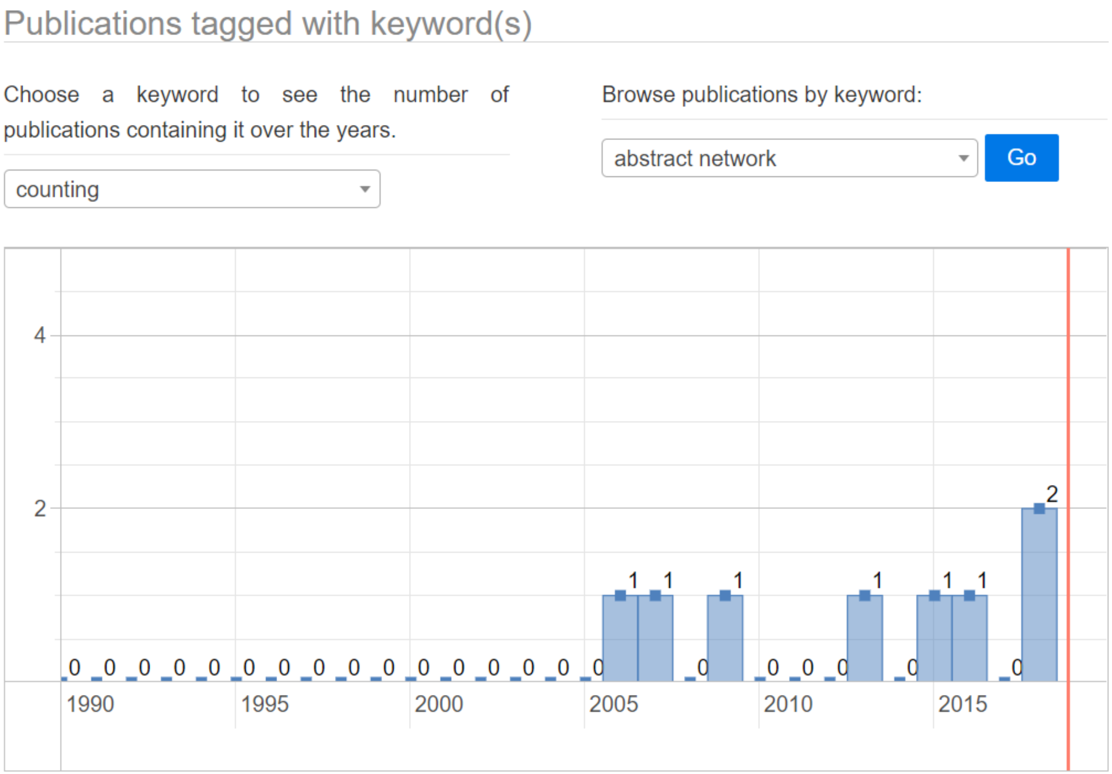
Bordewich & Semple, *Advances in Applied* Mathematics, 2016
Gunawan, DasGupta & Zhang, RECOMB 2017
Gunawan, AlCoB 2018
Weller, RECOMB-CG 2018

Fête de la Science at UPEM (science festival)

Photos : Campus numérique de l'UPEM

Outline

- An introduction to phylogenetic networks
- Classes of phylogenetic networks
- Counting problems on phylogenetic networks



Publications tagged with keyword(s)

Choose a keyword to see the number of publications containing it over the years.

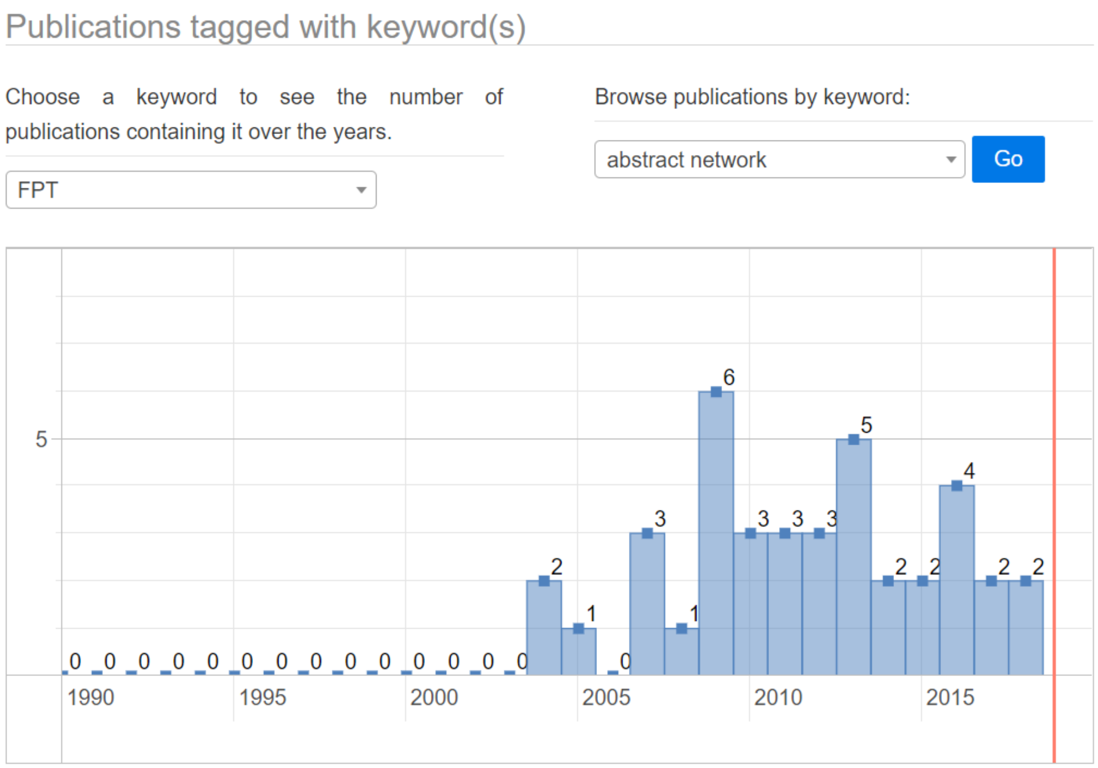
counting

Browse publications by keyword:

abstract network

Outline

- An introduction to phylogenetic networks
- Classes of phylogenetic networks
- Counting problems on phylogenetic networks



Publications tagged with keyword(s)

Choose a keyword to see the number of publications containing it over the years.

FPT

Browse publications by keyword:

abstract network

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.

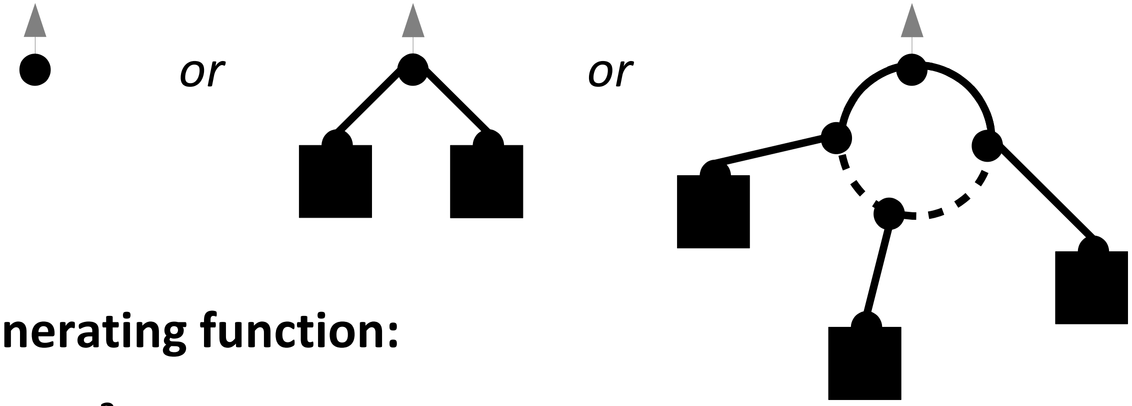
Semple & Steel, *TCBB*, 2006

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:

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

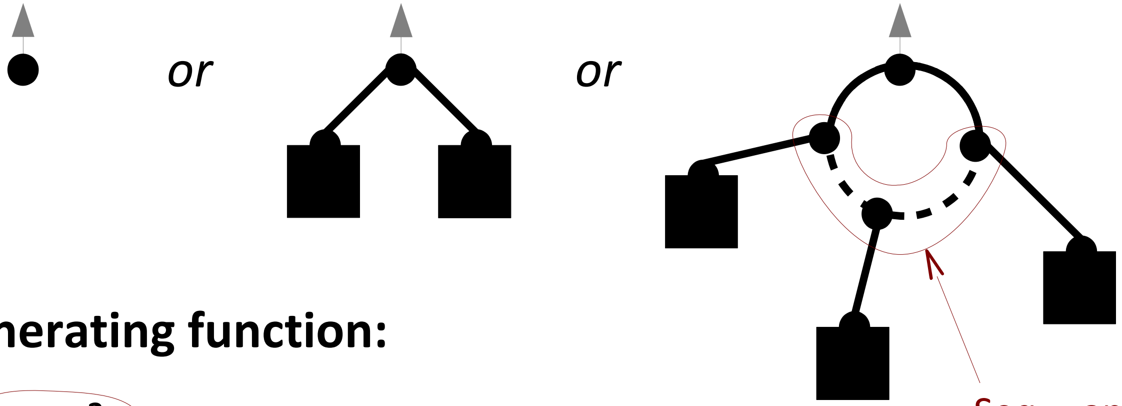
Semple & Steel, *TCBB*, 2006

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:

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

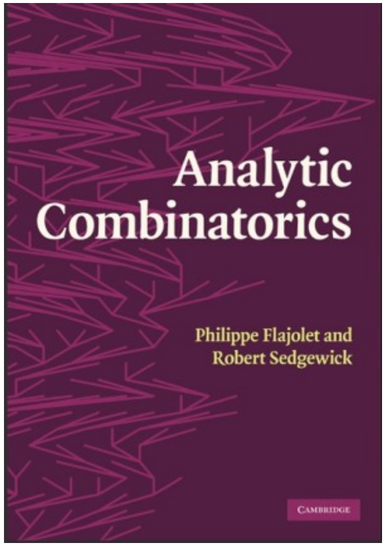
$\mathrm{Seq}_{\geq 2}$, any direction

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 ):

Analytic Combinatorics

Philippe Flajolet and Robert Sedgewick

Cambridge

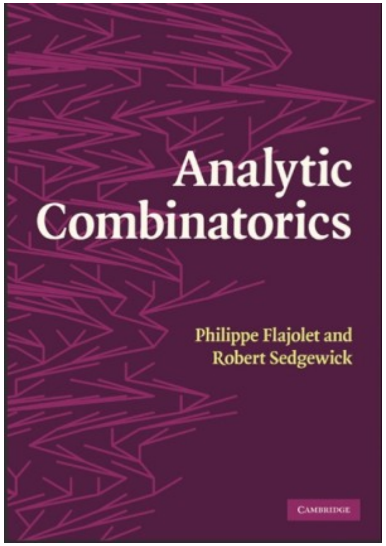
$$g_n \approx 0.2074\,(1.8904)^n\, n^{n-1}$$

Joint work with Mathilde Bouvel and Marefatollah Mansouri

Counting labeled unrooted level-1 networks

Exponential generating function:

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


Analytic Combinatorics
Philippe Flajolet and Robert Sedgewick
Cambridge

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 $\varphi(z)-z\varphi'(z)=0$

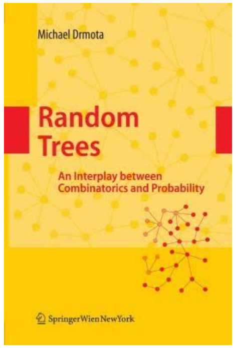
Joint work with Mathilde Bouvel and Marefatollah Mansouri

Counting labeled unrooted level-1 networks

Multivariate generating function:

$G(z,x,y) = \sum_{n,k,m} \frac{g_{n,k,m}}{n!} z^n x^k y^m$, where $g_{n,k,m}$ is the number of unrooted level-1 networks with $n+1$ leaves, k cycles, and m inner edges, satisfies

$$(\star_G) \qquad G(z,x,y) = z + \frac{1}{2}G(z,x,y)^2 + \frac{1}{2}xy^3\frac{G(z,x,y,)}{1-yG(z,x,y)}.$$

Using Theorem 2.23 of  :

Michael Drmota

Random Trees

An Interplay between Combinatorics and Probability

SpringerWienNewYork

Proposition 4.4. *Let X_n (resp. Y_n) be the random variable counting the number of cycles (resp. inner edges) in unrooted level-1 networks with $n+1$ leaves. Both X_n and Y_n are asymptotically normally distributed, and more precisely, we have*

$$\mathbb{E}X_n = \mu_X n + O(1), \qquad \mathbb{V}arX_n = \sigma_X^2 n + O(1) \quad \text{and} \quad \frac{X_n - \mathbb{E}X_n}{\sqrt{\mathbb{V}arX_n}} \xrightarrow{d} \mathcal{N}(0,1),$$

$$\mathbb{E}Y_n = \mu_Y n + O(1), \qquad \mathbb{V}arY_n = \sigma_Y^2 n + O(1) \quad \text{and} \quad \frac{Y_n - \mathbb{E}Y_n}{\sqrt{\mathbb{V}arY_n}} \xrightarrow{d} \mathcal{N}(0,1),$$

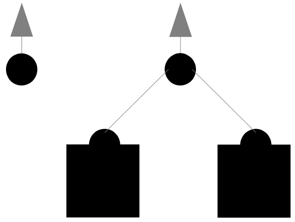
where $\mu_X \approx 0.46$, $\sigma_X^2 \approx 0.18$, $\mu_Y \approx 1.61$ and $\sigma_Y^2 \approx 1.44$.

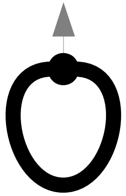
Joint work with Mathilde Bouvel and Marefatollah Mansouri

Counting labeled unrooted level-2 networks

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

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





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

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

→ Seq_{1}

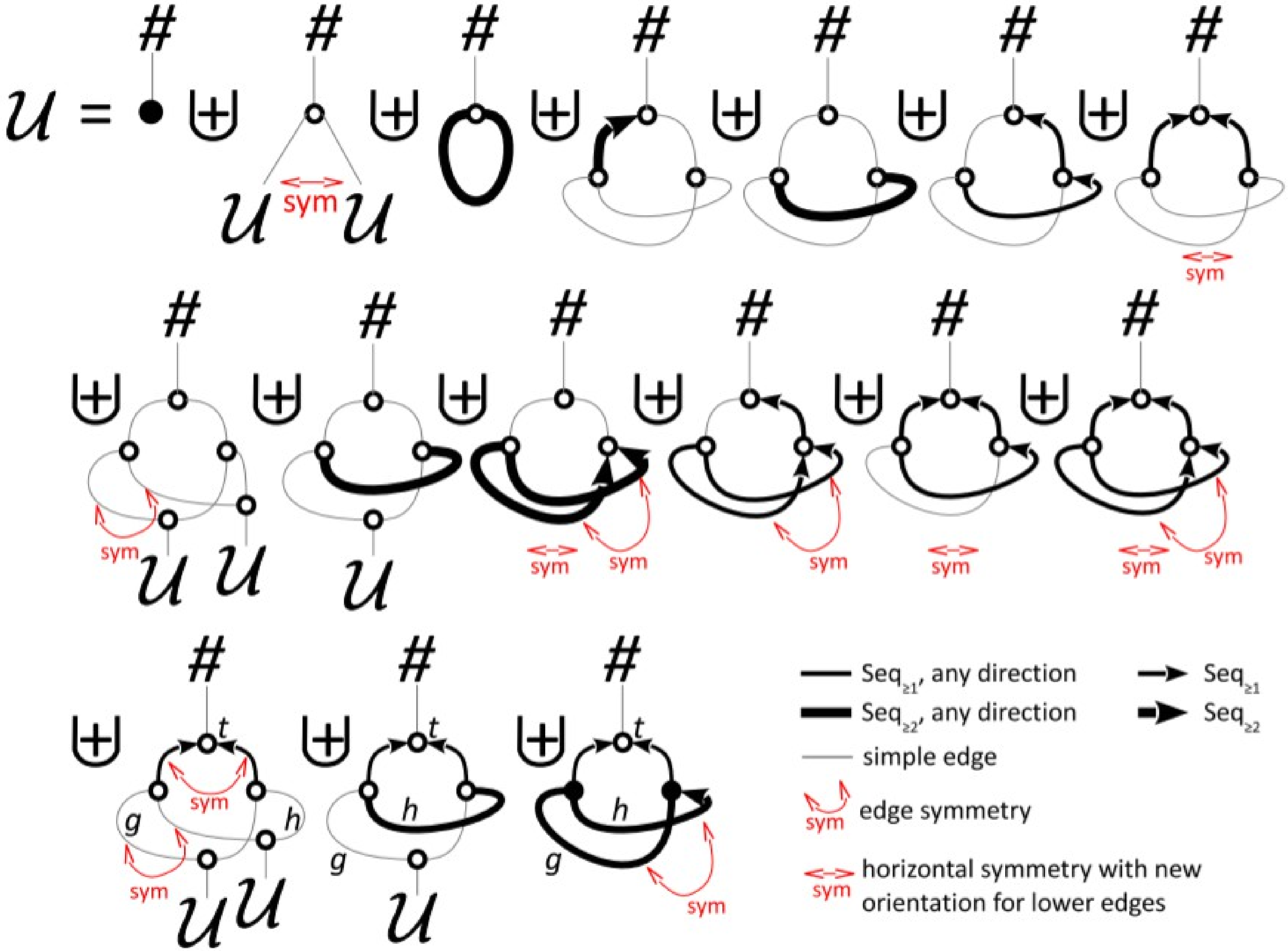
➡ $\text{Seq}_{\geq 2}$

— simple edge

sym edge symmetry

<-> sym horizontal symmetry with new orientation for lower edges

Counting labeled unrooted level-2 networks



Joint work with Mathilde Bouvel and Marefatollah Mansouri

Counting labeled unrooted level-2 networks

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

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

Rewrite:

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

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:

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

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:

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

Lagrange inversion:

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

Taylor expansions + Newton formula

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

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 :

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

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

Counting labeled level-*k* networks

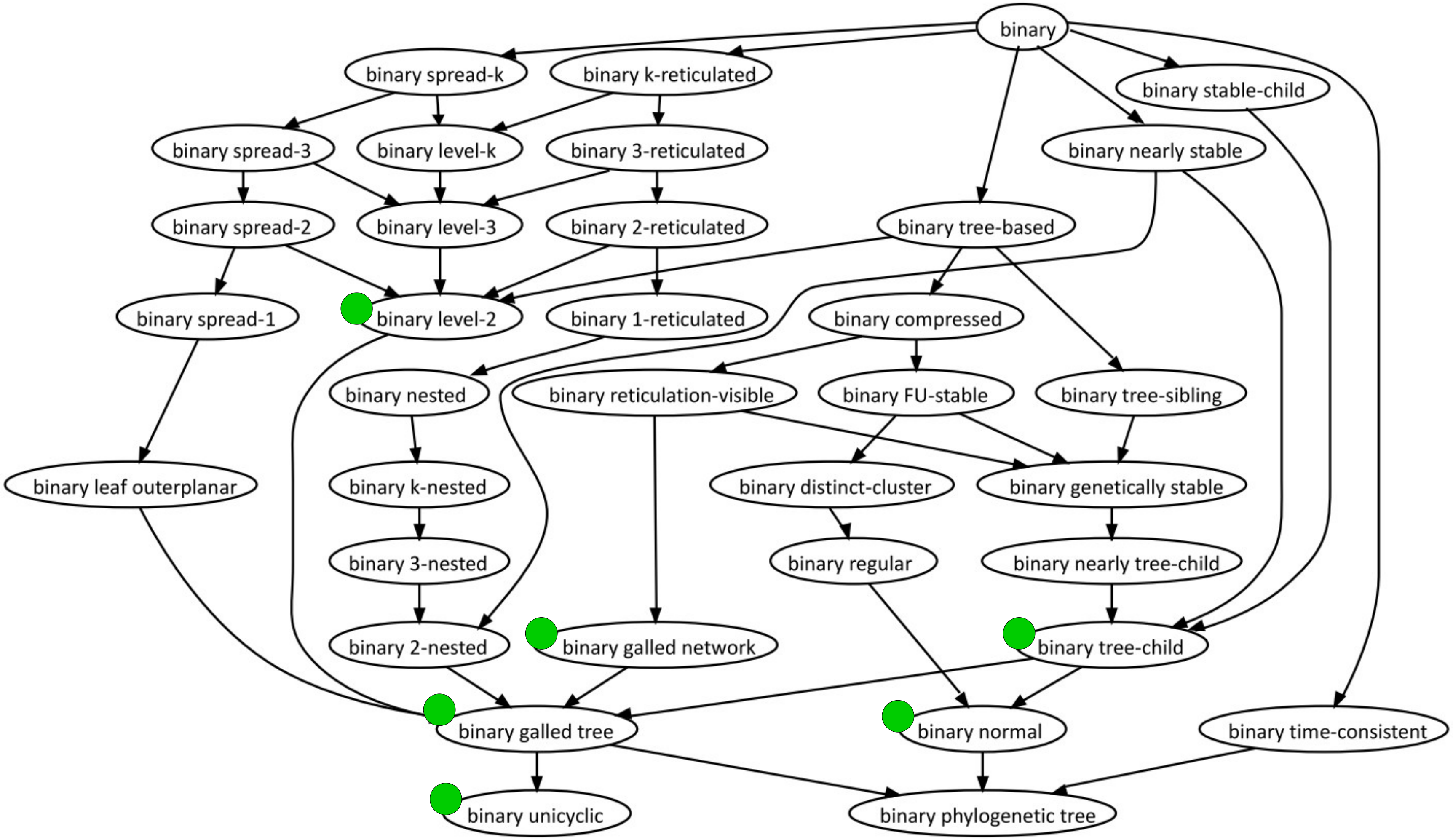
Open problem on level-1 networks = galled trees:
Counting galled tree "shapes" (= unlabeled networks)

Upper bound

Chang, Hon & Thankachan, Data Compression Conference 2018

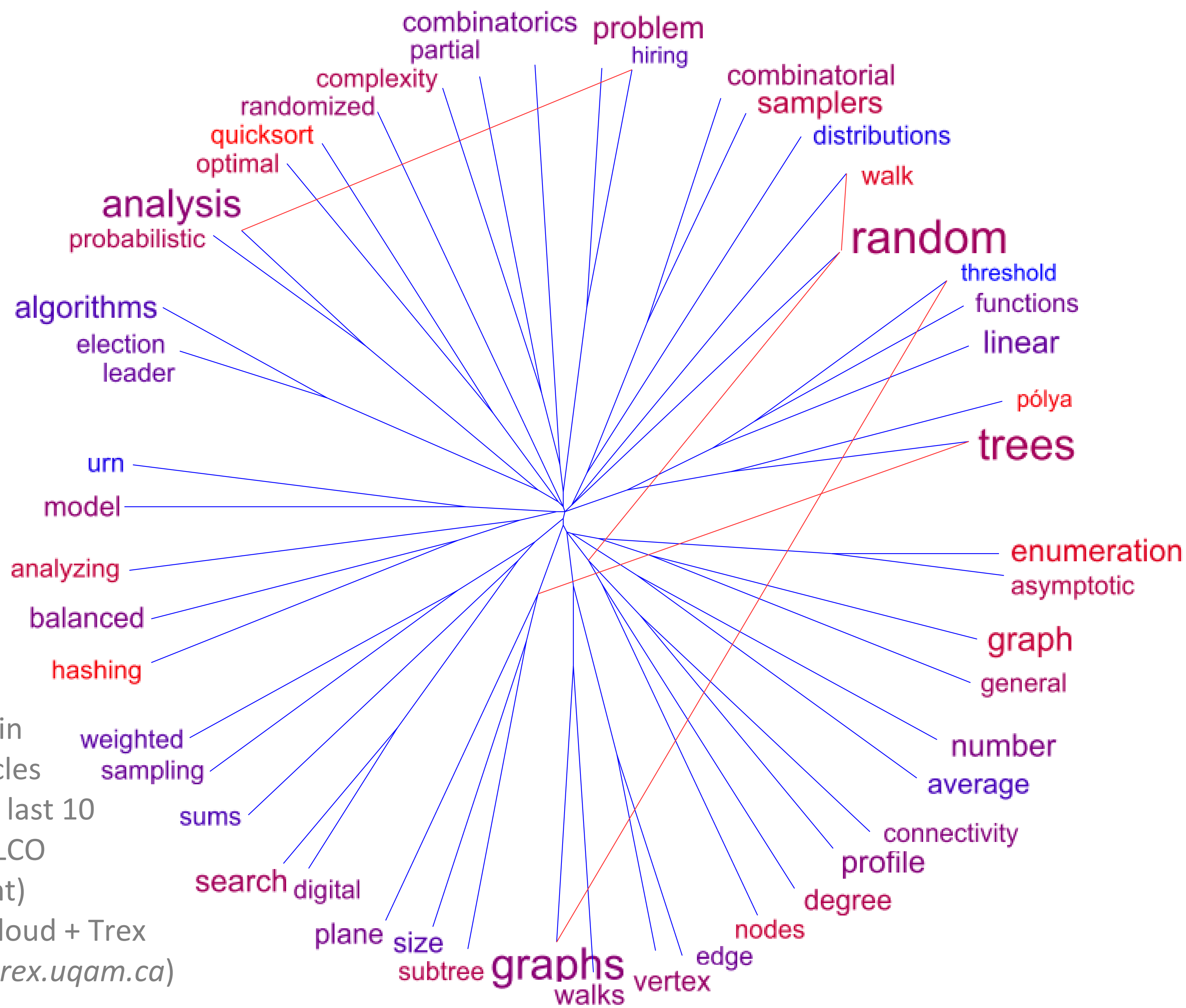
Upper levels...

Other phylogenetic network counting results



Tree-child, normal networks: McDiarmid, Semple & Welsh 2015
Fuchs, Gittenberger & Mansouri 2018
Galled networks: Louxin Zhang 2018

Thank you for your attention!



Reticulogram
of the 100 most
frequent words in
the titles of articles
published in the last 10
editions of ANALCO
(red: most recent)
Built with TreeCloud + Trex
(*treecloud.org*, *trex.uqam.ca*)