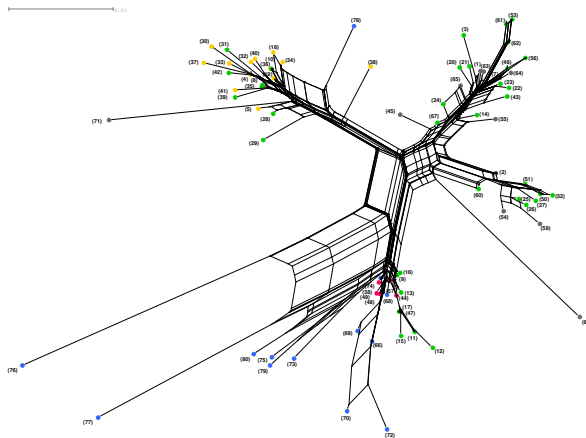


Visualizing reticulate evolution with planar split networks

Andreas Spillner

Cambridge June 2011

Outer-labeled split networks



- ▶ Network generated in SplitsTree using NeighborNet
- ▶ 80 mitochondrial DNA-sequences used in a study of geographic range expansions in a species of oak gallwasp [Stone et al. 2001, 2007]

From sequence data to split network

A	C	C	A	T	A	T	C	C	G
A	T	C	A	T	A	T	C	C	G
A	T	C	A	T	A	T	C	T	G
A	T	C	A	T	A	T	C	T	G
A	C	C	A	T	A	T	C	T	G

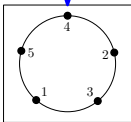
sequence alignment

distance transformation

0	1	2	2	1
1	0	1	1	2
2	1	0	0	1
2	1	0	0	1
1	2	1	1	0

distance matrix

agglomeration



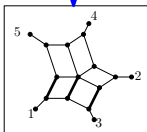
circular ordering

extracting/weighting splits

0.0023	1, 3 2, 4, 5
0.1203	2, 4 1, 3, 5
0.0106	3 1, 2, 4, 5
...	

set of weighted splits

draw network

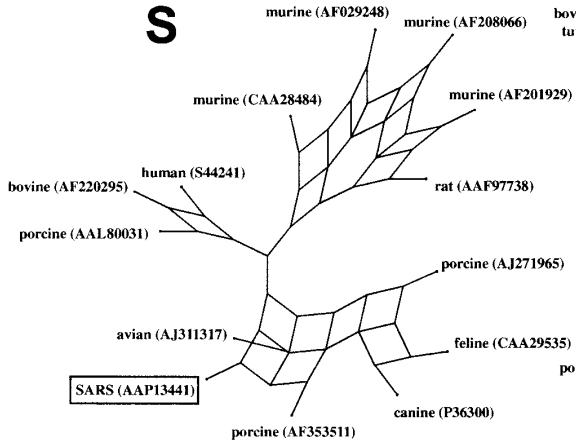


outer-labeled planar split network

NeighborNet
[Bryant and Moulton, 2004]

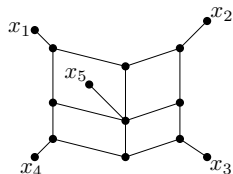
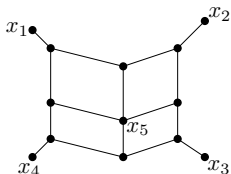
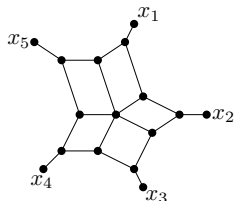
Drawing algorithm
[Dress and Tsuru, 2004]

Non-outerlabeled split networks



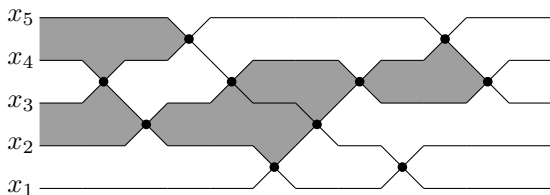
- ▶ Network generated using SplitDecomposition presented in a study of the SARS coronavirus [Stavrínides and Guttman, 2004]

Planar split networks and flat split systems



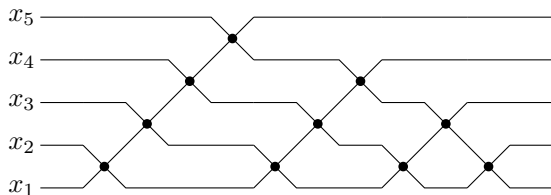
- ▶ Flat split systems were originally defined in terms of oriented matroids of rank 3. [Bryant and Dress, 2007]
- ▶ Intuitively: a split system Σ is *flat* if it can be displayed by a planar split network
- ▶ Open: Is there a polynomial time algorithm to decide whether a split system is flat?

You are flat? Make sure you can prove it to me!



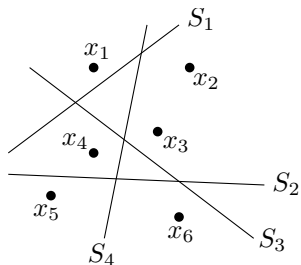
- ▶ Σ flat — splits in Σ correspond to faces in a wiring diagram.

Okay, but where do I get the wiring diagram from?



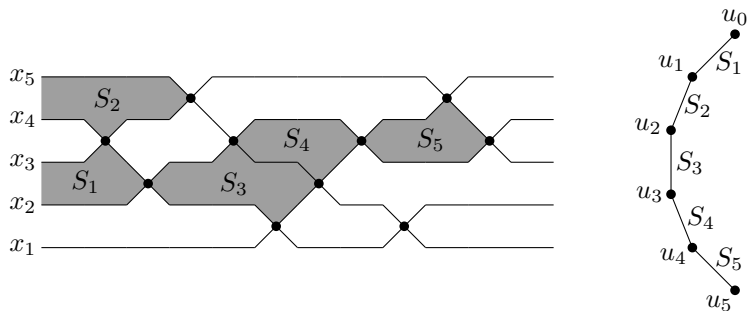
- ▶ For any circular split system one can easily construct a suitable wiring diagram.

And we can also get “non-circular” wiring diagrams



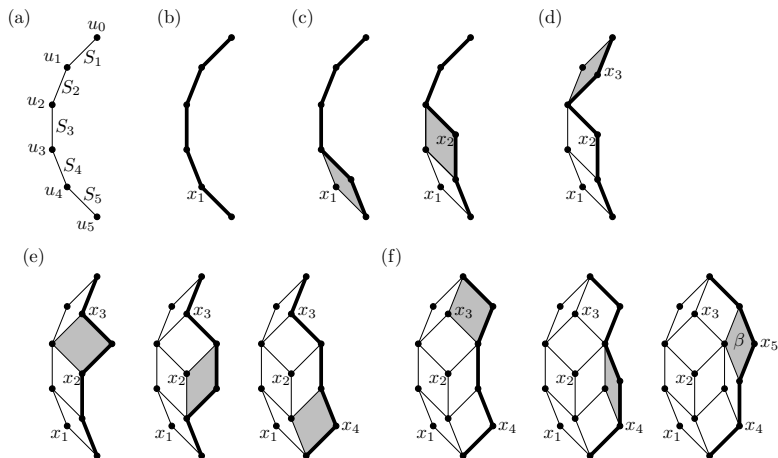
- ▶ Σ affine — map X to a set of points in the plane that need not be in convex position
- ▶ point $p = (a, b) \rightarrow$ straight line $y = ax - b$
- ▶ This could be interesting in phylogeographic applications where geographic coordinates of sampling locations are often known.

Transforming the wiring diagram into a split network

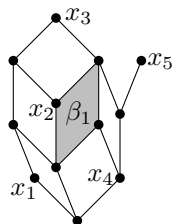
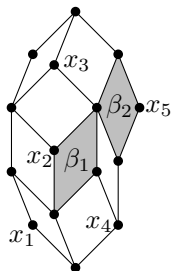


- ▶ Approach is similar to computing the dual of an arrangement of pseudolines [Agarwal and Sharir, 2005]
- ▶ We maintain a path P in the network constructed so far.

Processing the labels from bottom to top



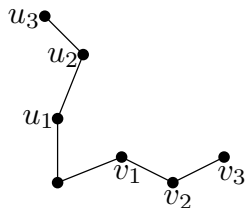
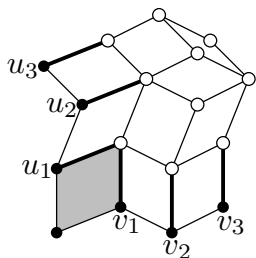
Do we need all those boxes?



- ▶ Box β_1 conveys important information about the splits.
- ▶ Box β_2 is not needed — the two corresponding splits $S = A|B$ and $S' = A'|B'$ are *compatible*, that is, at least one of the intersections is empty:

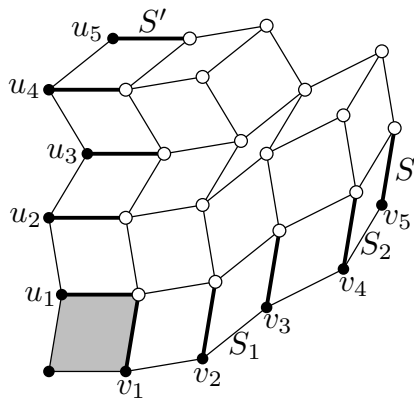
$$A \cap A', \quad A \cap B', \quad B \cap A', \quad B \cap B'$$

Just cut it off!



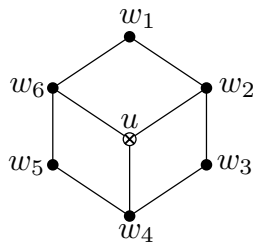
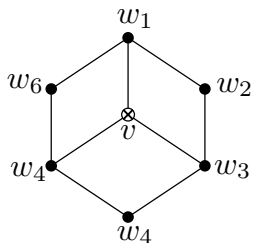
- Sometimes the box (and even some more vertices) can easily be removed.

And if this doesn't work?



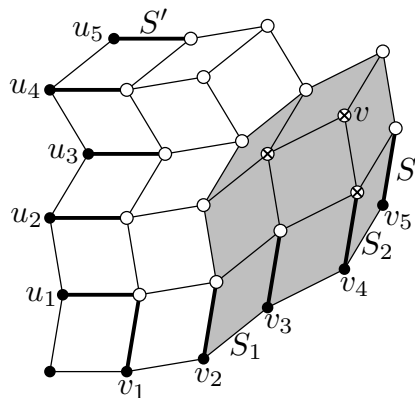
- There might be splits that share a box with both S and S' .

Modify the network locally



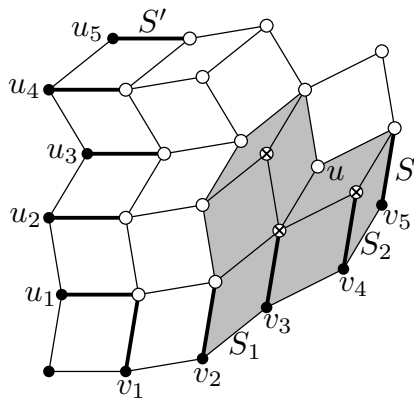
- Perform a flip around a vertex.

Find a sector



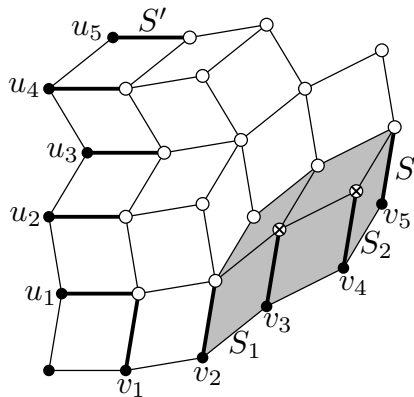
- In the sector we can always perform at least one flip.

Clear the sector



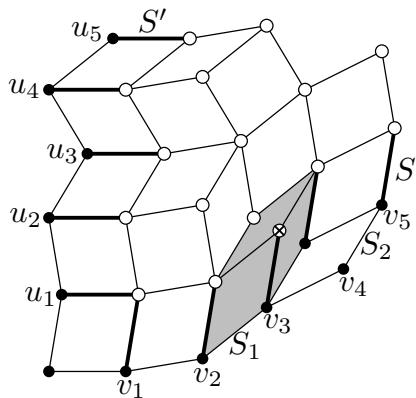
- Keep on flipping...

Clear the sector



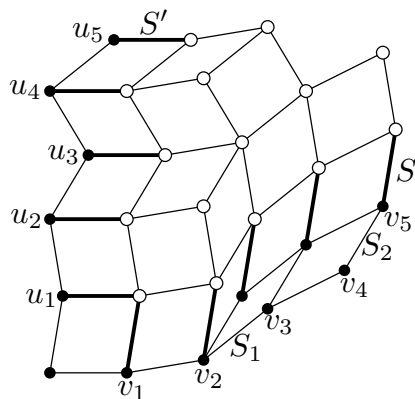
- Keep on flipping...

Clear the sector



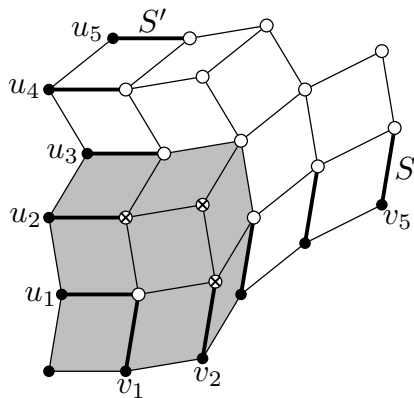
- Keep on flipping...

Clear the sector



- ▶ ...until the sector disappeared completely.
- ▶ Then the number of “white” vertices is strictly smaller than before.

And then?

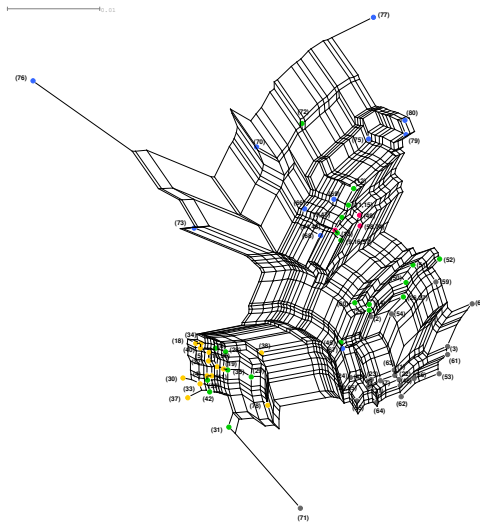


- ▶ Either we can just cut it off...
- ▶ ...or we find and clear another sector.

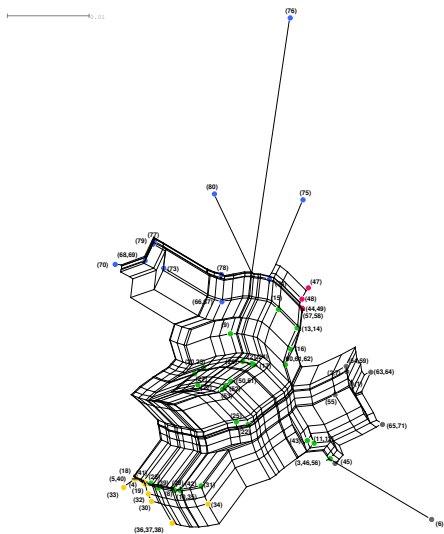
Run time

- ▶ Given m splits on a set X with n elements represented by an allowable sequence.
- ▶ The initial split network can be computed in $O(mn + m^2)$ time.
- ▶ The transformation into a minimal network can be done in $O(nm^2 + m^6)$ time.
- ▶ The network has $O(m^2)$ vertices and edges.

Gall wasp data set: Flat split systems via multidimensional scaling



Gall wasp data set: Flat split systems from geographic locations



Collaborators and Acknowledgments

Vincent Moulton



Binh Nguyen



Acknowledgments:

British Council and DAAD (ARC Programme)

Thank you