Understanding NeighborNet:

> Simon J. Greenhill (simon@simon.net.nz).
>
> ARC Centre of Excellence for the Dynamics of Language, Australian National University.
>
> Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History.

1. A tree

Here is some simple data – with no conflicting signal, so the tree fits perfectly.

	A	B	C	D	E	F	G	H	I
language1	1	0	0	0	0	1	1	0	0
language2	0	1	0	0	0	1	1	0	0
language3	0	0	1	0	0	0	0	0	1
language4	0	0	0	1	0	0	0	1	1
language5	0	0	0	0	1	0	0	1	1

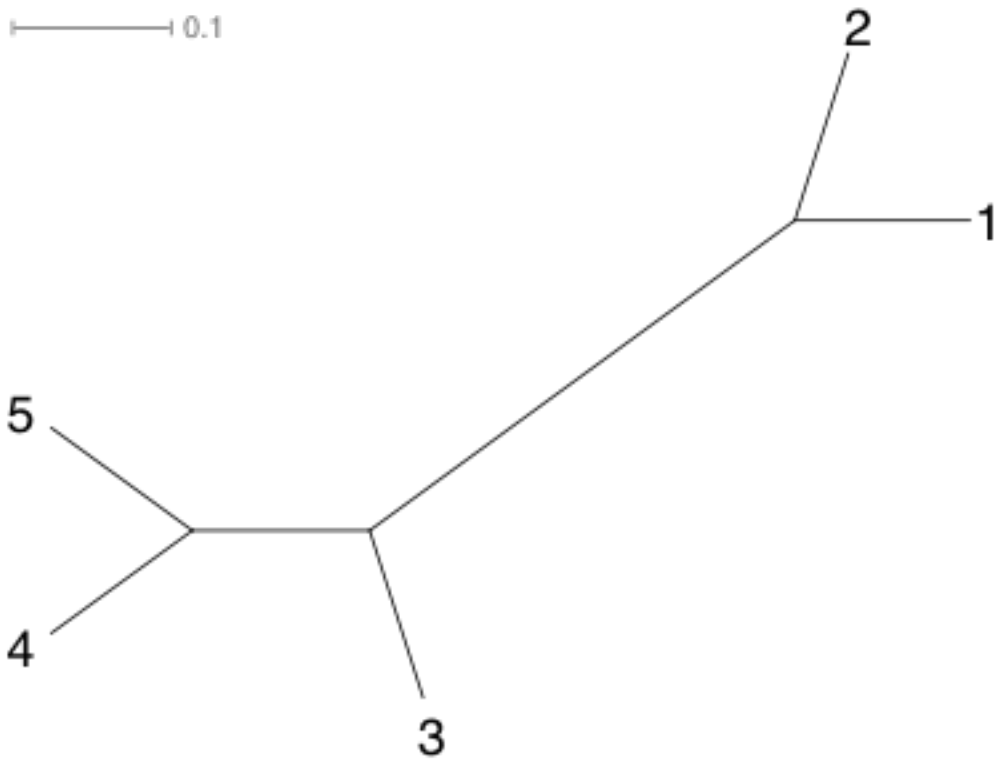
The first 5 characters (`A-E`) are singletons – i.e. are a cognate set present in one language alone. Ignore them – they're just here to give the tree some branch-lengths.

The characters `F` and `G` group language 1 and 2 together.

Character `H` groups languages 4 and 5 together.

Character `I` groups language 3, 4 and 5.

This is what the network looks like – it's a tree:



Also see the scale bar in the top left corner. There are 9 characters in our data, so you can see that it is proportional to the data in our matrix. The branch between 4,5 and 3 is supported by 1 character – and is about 1/9 in length.

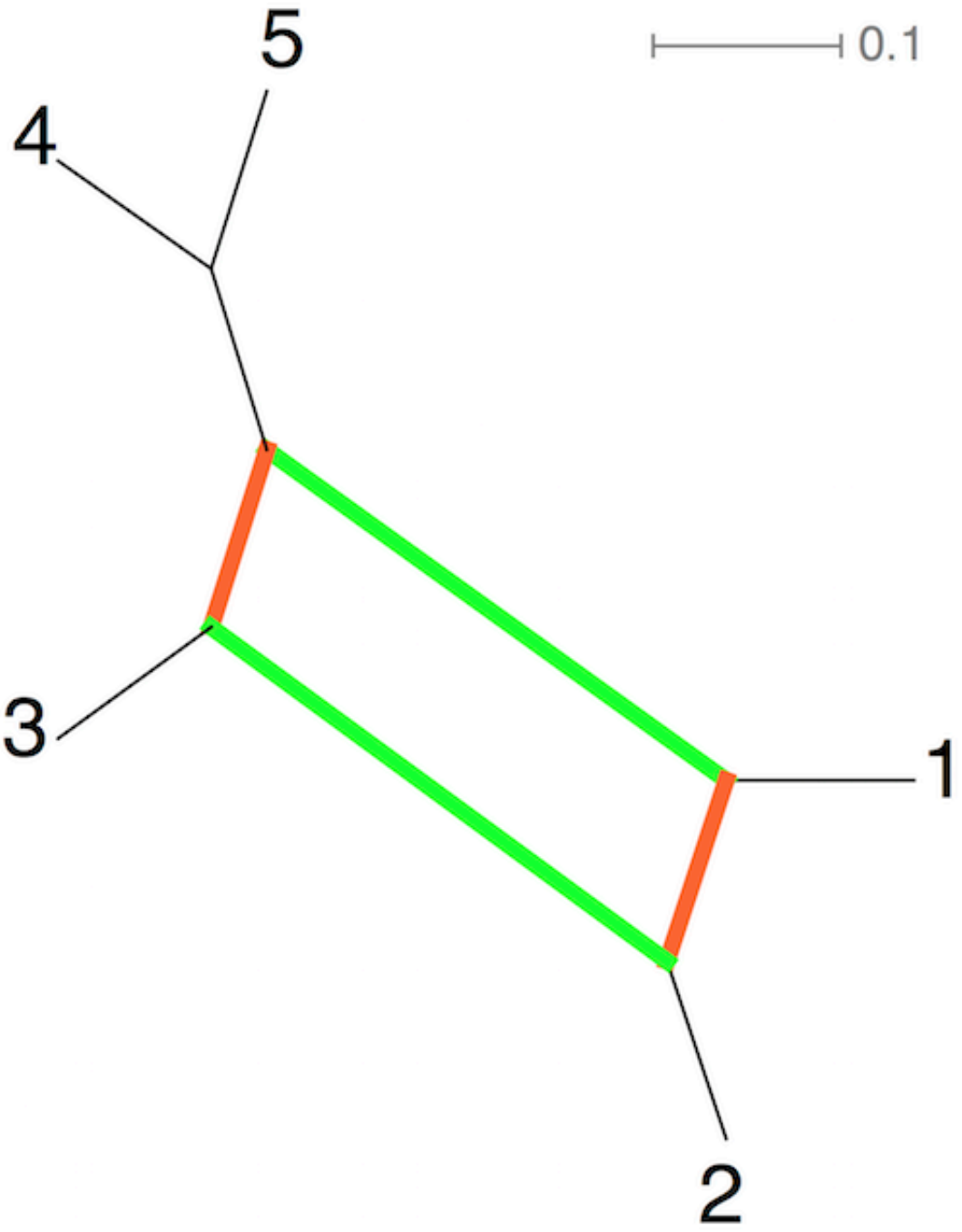
2. A simple network.

We've added one extra character to the data – `J` – this groups languages 2 and 3 together – in conflict with the groupings we saw above.

	A	B	C	D	E	F	G	H	I	J
language1	1	0	0	0	0	1	1	0	0	0
language2	0	1	0	0	0	1	1	0	0	1
language3	0	0	1	0	0	0	0	0	1	1
language4	0	0	0	1	0	0	0	1	1	0
language5	0	0	0	0	1	0	0	1	1	0

See that the groupings (4,5), (3,4,5) and (1,2) are the same as the above tree.

However, now that we have conflicting signal, we get a network:



Note the box representing the conflicting signal that we added in character `J`. The trick is to read across parallel lines. In the below neighbor-net we have two sets of parallel lines: one in green, and one in orange.

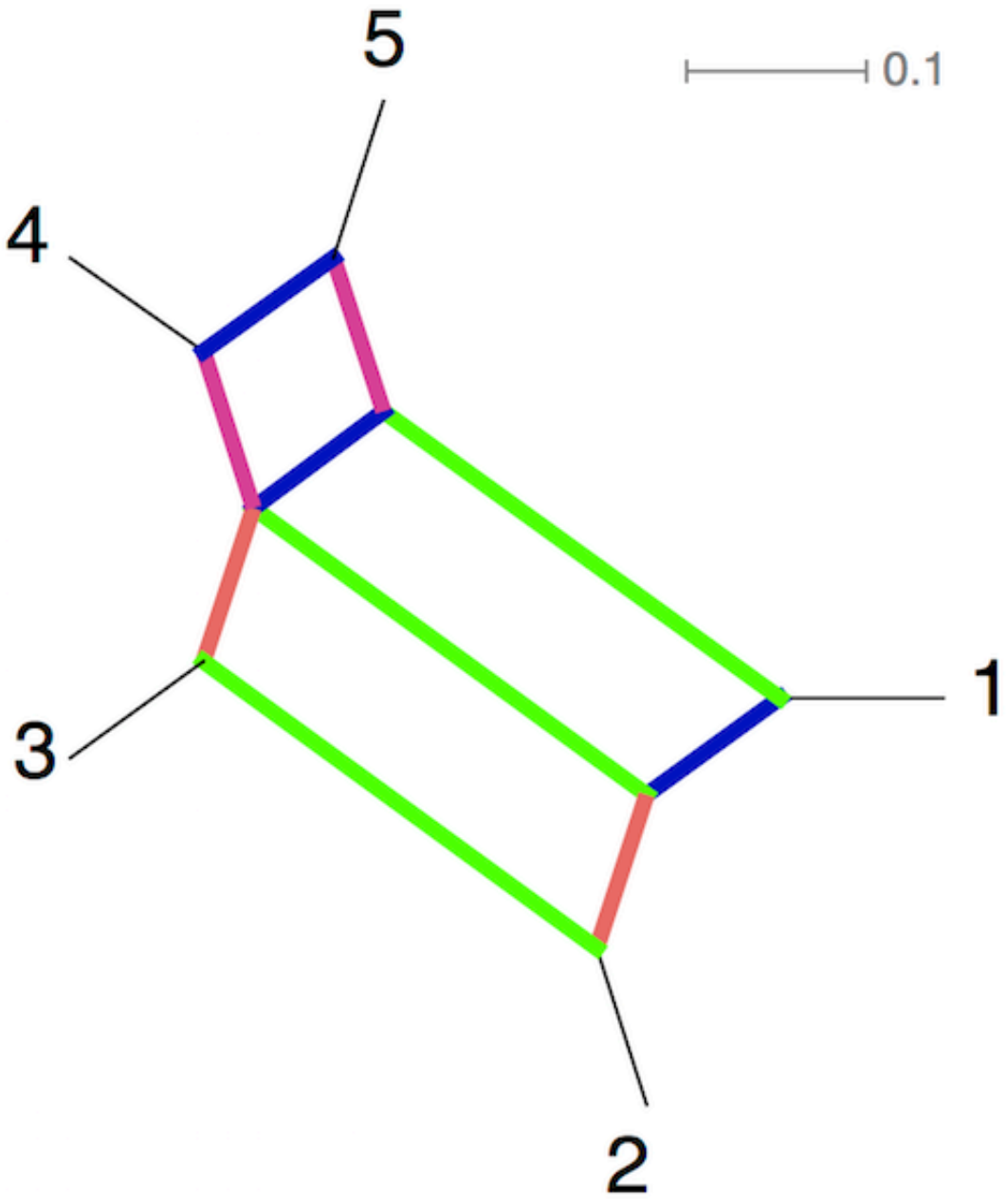
We have 10 characters now, and 1/10 characters (i.e. `J`) says that 2 and 3 group together. This is the short edge of the box below (in orange), which measures about 1/10th in length. The other side of the box (green) is longer – about 2/10ths – which is the 2 character (`F` and `G`) that put language 1 and 2 together.

3. A more complex network

Now we've added character `K`. This adds more conflict grouping languages 1 & 5 together.

	A	B	C	D	E	F	G	H	I	J	K
language1	1	0	0	0	0	1	1	0	0	0	1
language2	0	1	0	0	0	1	1	0	0	1	0
language3	0	0	1	0	0	0	0	0	1	1	0
language4	0	0	0	1	0	0	0	1	1	0	0
language5	0	0	0	0	1	0	0	1	1	0	1

See that we have mostly the same groupings as before. The conflicting character `K` has added a box between (1,5) and (2,3,4) – in blue. This also means that the character H is conflicting because (1,5) conflicts with (4,5). Therefore the lines in purple get added. Finally, an extra green line is added to connect everything up.


5
0.1
4
1
3
2