EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

UPPSALA
UNIVERSITET

Phylogenetic inference of reticulate evolution in Tupí-Guaraní languages

Fabrício F. Gerardi (UT) & Tiago Tresoldi (UU)

ICHL25
Oxford, 2022/08/04

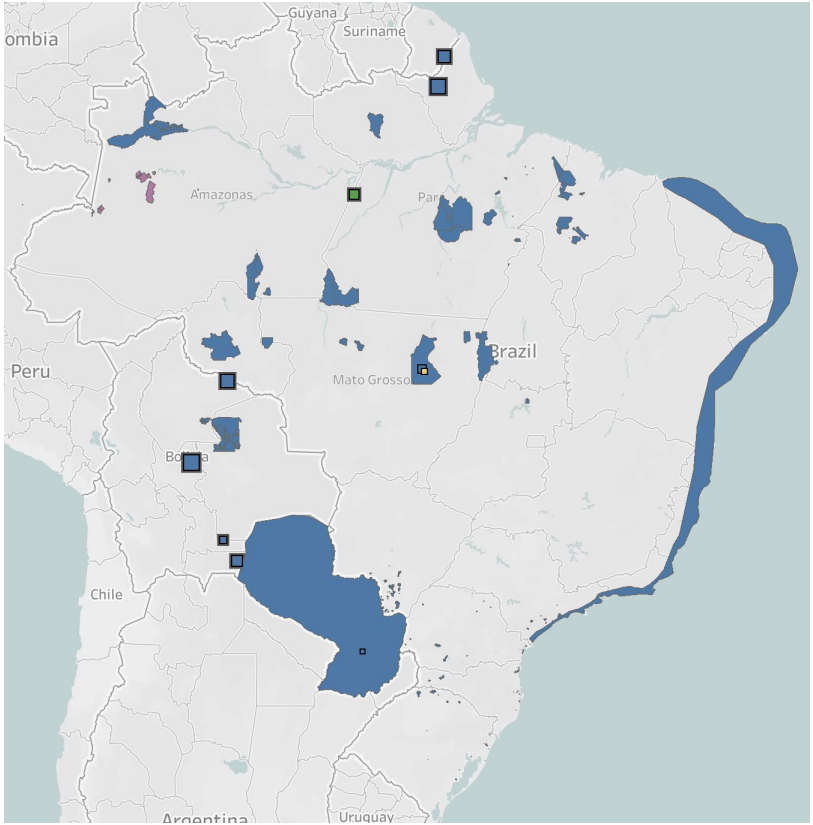
Outline

- Introduction
  - Tupí-Guaraní languages and archeology
  - Reticulation
- Material and methods
  - Tupí-Guaraní tree
  - Experiments with method for detecting reticulation
- Ongoing and future work

Tupí-Guaraní

- Largest branch of the Tupian family
  - Over 40 languages
  - Spreading 4000 km in latitude and in longitude
  - Languages with less than 50 speakers (e.g. Avá Canoeiro) and languages with millions of speakers (e.g. Avañe'e)
- Lack of consensus on the date of the root, the geographic origin, and the migration
  - Gerardi et al. (2022) has our proposed tree
- History marked by events of contact, borrowing, and hybridization/admixture, as most native American families

Tupí-Guaraní Languages



Tupí-Guaraní languages (blue), Mawé (green), Awetí (Yellow), Omagua and Kokama (pink).

Competing theories

Two main migration patterns

- Fluvial network promoting a rapid expansion; war-like ideology
- Population increase as a key driver, floodplain agriculture, slower movements (Iriarte et al. 2017)

Two main migration patterns

- South to North-East
- East to South-West

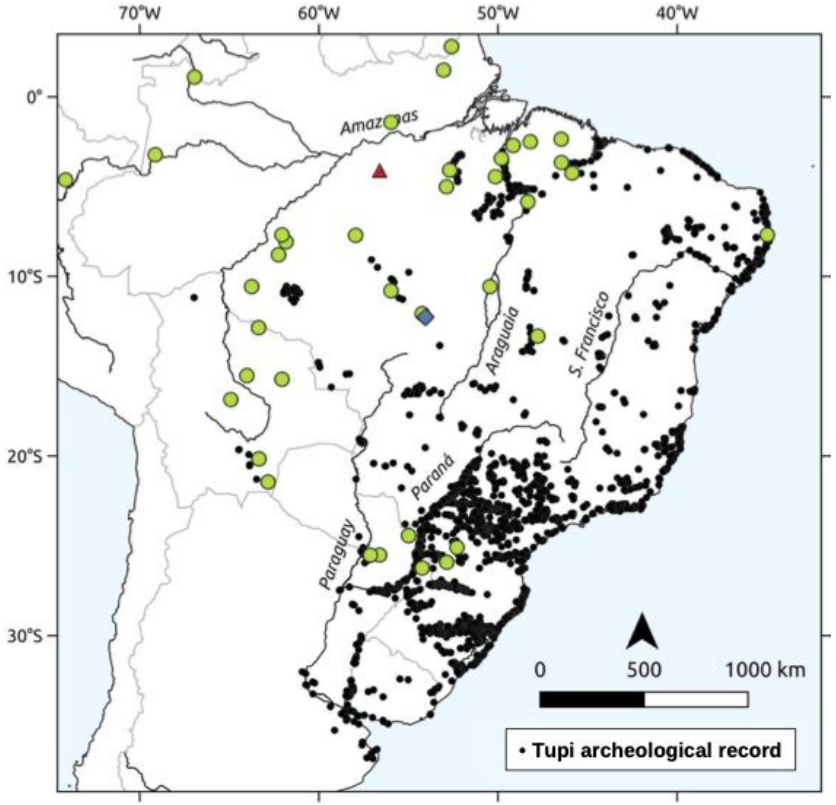
Archaeological evidence

The dispersal of TG languages has a clear material correlate (“Tupiguarani tradition”) in the spread throughout eastern South America of a package including a particular type of ceramics, plant management, and cultivation of a variety of crops (Balée 2000, 2006, 2014; Brochado 1984; Noelli 1998, 1999; Noelli et al. 2008).

C14 Chronologies

- Eastern Amazon: 1670 BP (+/- 80)
- Bolivia: 1680 BP (+/- 90)
- Northeast Brazil: 1690 BP (+/- 110)
- Paraná Basin: 2010 BP (+/- 75)
- Atlantic coast: 1055 BP (+/- 80)

Archaeology and Language



- Archaeological dates considered too ancient have often been discarded, based on the view that the TG dispersal is a recent one.
- Over time the accumulation of dates close to ca. 2000 BP in different regions led to a questioning of this premise.
- Glottochronological estimates of ca. 2500 BP for the initial split of the TG languages (Rorigues 1964, 2000a,b, 2005; Urban 1996)

Reticulation

- Human history is marked by events of reticulation, linguistics also
- Trees are a narrative device that can suggest evolutionary paths masking too much of the complexities
- Even if we can obtain a “perfect” tree, correctly mapping all vertical transmissions, it might only partially represent the history we want to tell
- Reticulation is an important step for bringing to light the history of native American families

Data

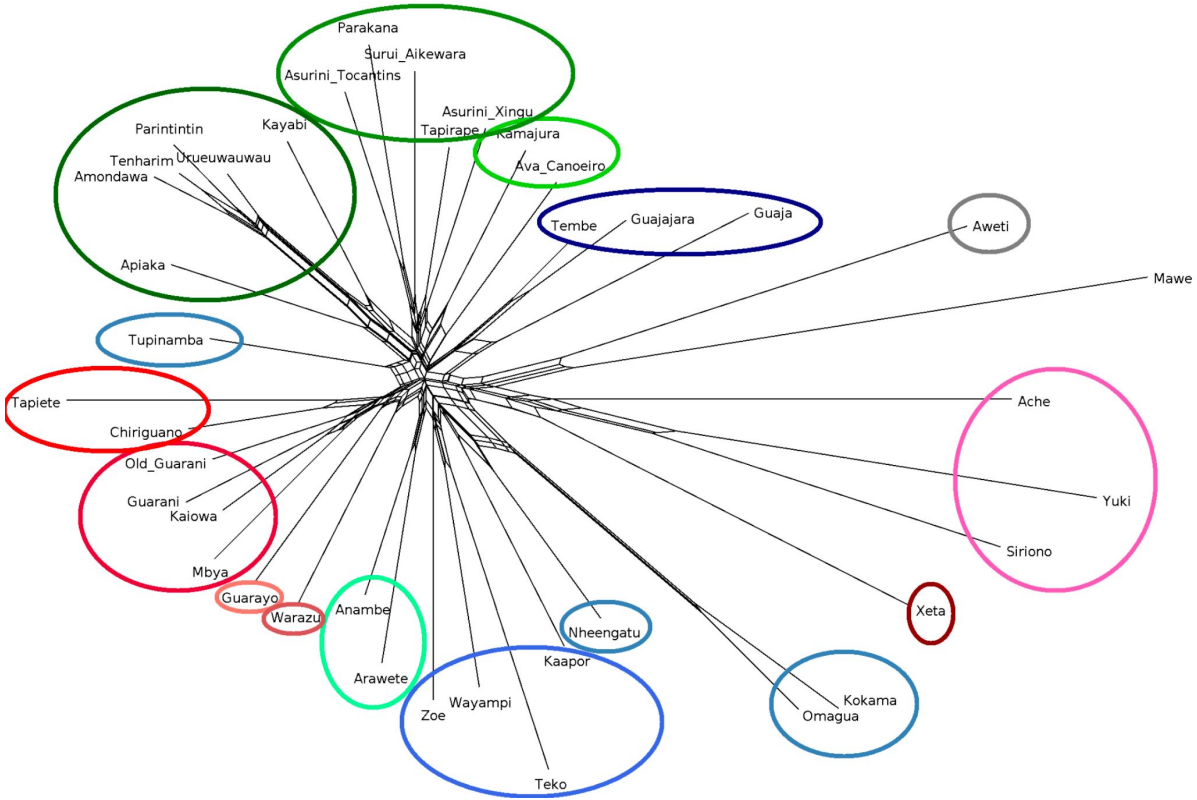
Data from TuLeD (Tupían Lexical Database) (Gerardi et al. 2022)

- **Varieties:** 40 (38 TG, 2 Tupi)
- **Concepts:** 447 (improvements to appear in next release)

Language	Concept	Phonetic form	Cognate set
Tupinambá	BAT	anira	171
Wayampi	BAT	aniɬa	171
Guaraní	BAT	mopi	172
Kaiowá	BAT	ᵐbopiri	172
Mawé	BAT	haki?i	4513

- https://tular.clld.org

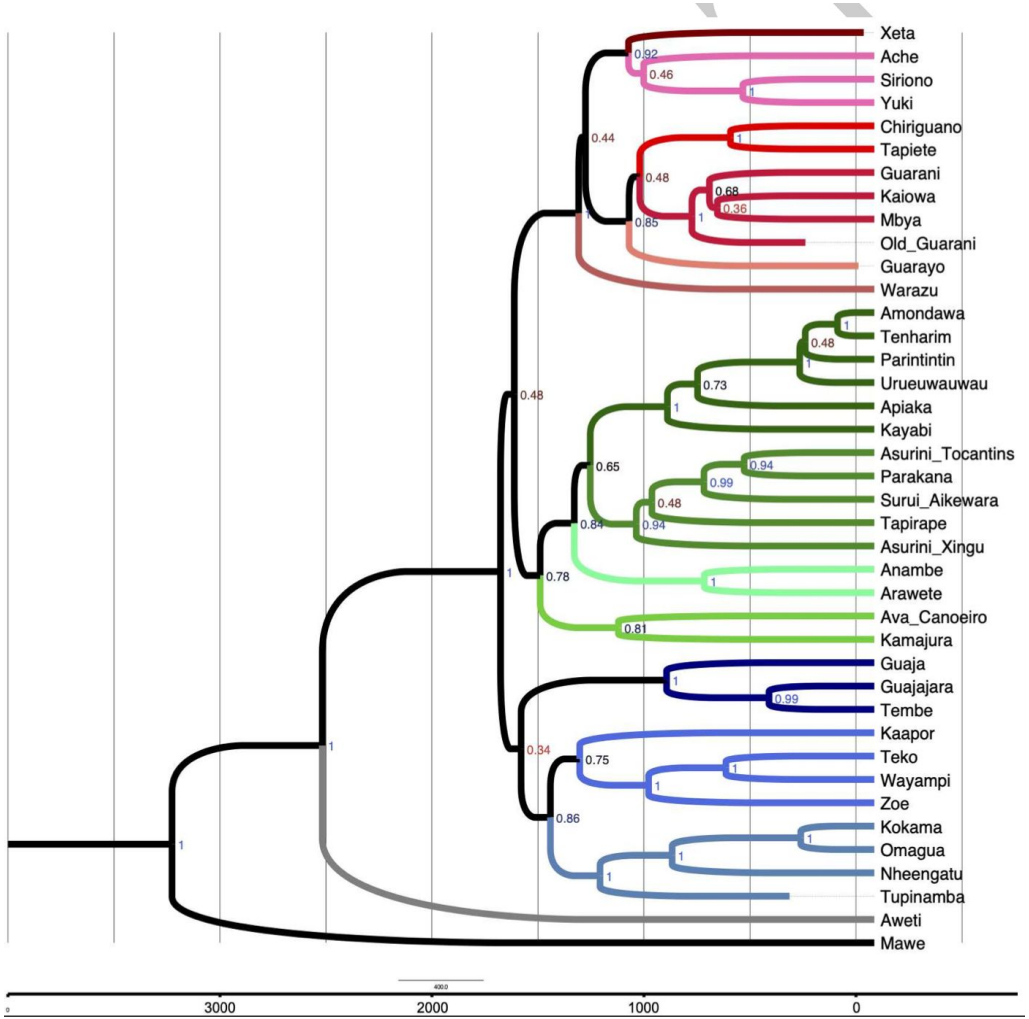
Neighbor-Net



- The network shows a level of hybridization similar to that found in other families (Kolipakam et al. 2018)
- The position of Tupinambá (far from its known descendant Nheengatu) is likely motivated by the high level of borrowings to most of the tree.
- Some of the hybridizations suggested are not compatible with the known history, and might be due to the borrowings (Schleicher 1998).

Classification

- Bayesian inference with BEAST2 (Bouckaert et al. 2019)
  - Binary Covarion model
  - Birth-Death tree
  - Relaxed lognormal clock
- Experimented with different models
- Results and models available as supplements of Gerardi et al. (2022)

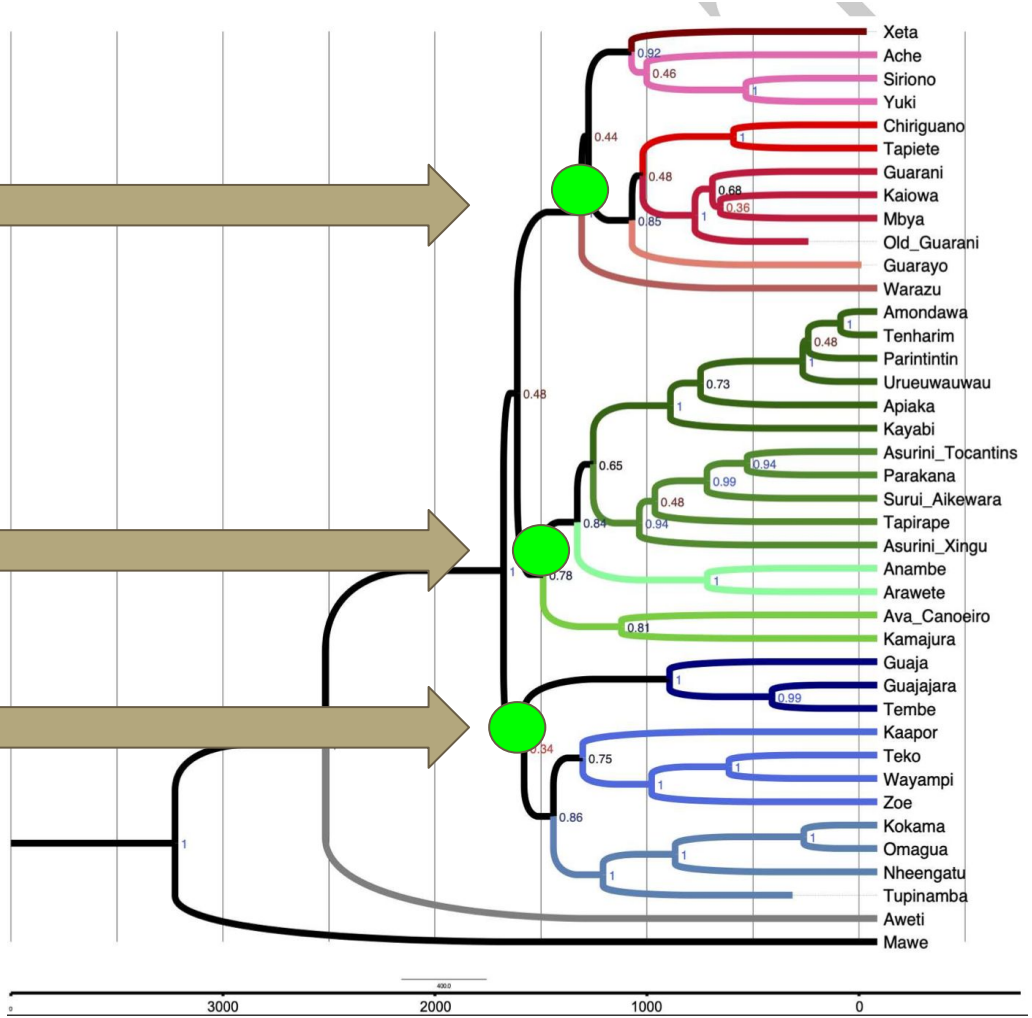


Classification

Southern group

Core group

Northern-and-Eastern group



Reticulations

- A high level of reticulation in TG has been argued since the earliest classification studies (Schleicher 1998; Mello 2000).
  - Supported by a growing body of comparative linguistic and anthropological data, and historical evidence.
  - Such events, more intensive than mere borrowings, are probably the cause for long-standing uncertainties in the position of languages such as Avá-Canoeiro, Kamajurá, and Tenetehara.
- This evidence aligns the case of Tupí-Guaraní with the ongoing discussion on how suitable trees are for narrating linguistic histories (cf. de Castro 1984, 1996; Noelli 1996), and provides a good case for testing methods for detecting reticulation.

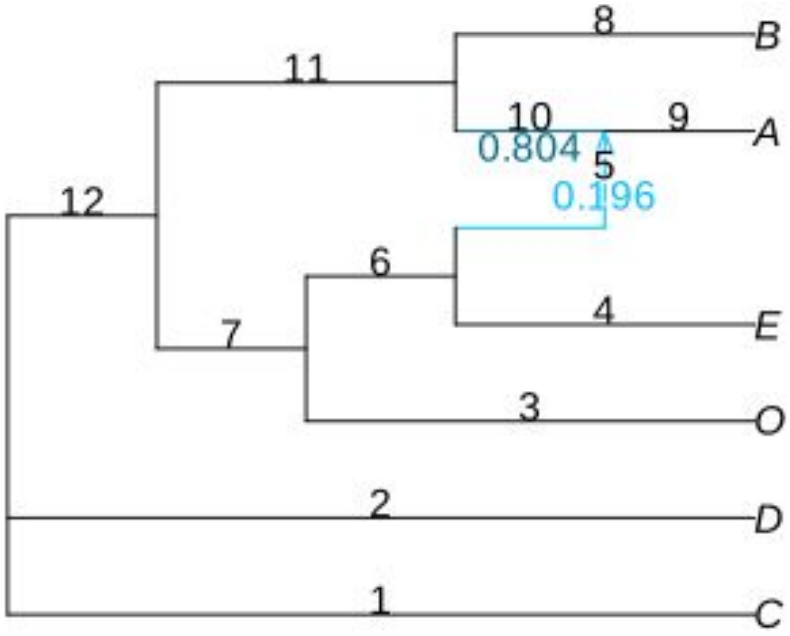
Networks

- Most classifications maintain the tree model as their goal (Rodrigues 1984; Mello, 2001, Cabral & Rodrigues 2002; Michael et. al 2013, Gerardi & Reichert 2021), leading to a theoretical debate influenced by the rising prominence of networks in phylogenetics
- We argue that, **once the major traits of vertical inheritance have been detected, networks can be more suitable for the study of reticulate phylogenies**, better reflecting hybridization and contact.
- The reconstruction of such evolution requires adopting a set of practices and tools developed for biological classification still under-exploited in computational historical linguistics.

Methods - I

We first tested the method by Solís-Lemus & Ané (2016), *PhyloNetworks.jl / SNAQ*, unable to use it for our linguistic data

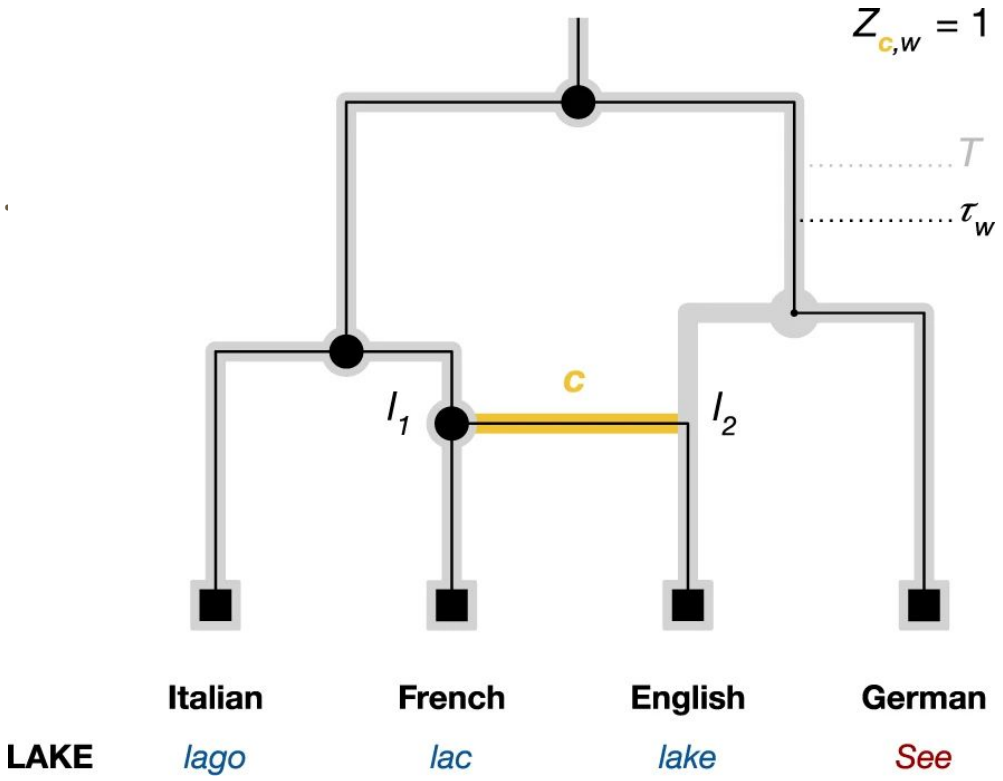
- Genetic vs. binary data
- *Long* processing time
- Only providing the percentage of inferred reticulation (actual characters need to be detected with other methods)



Methods - II

We tested the contacTrees method recently published by Neureiter at al. (2022)

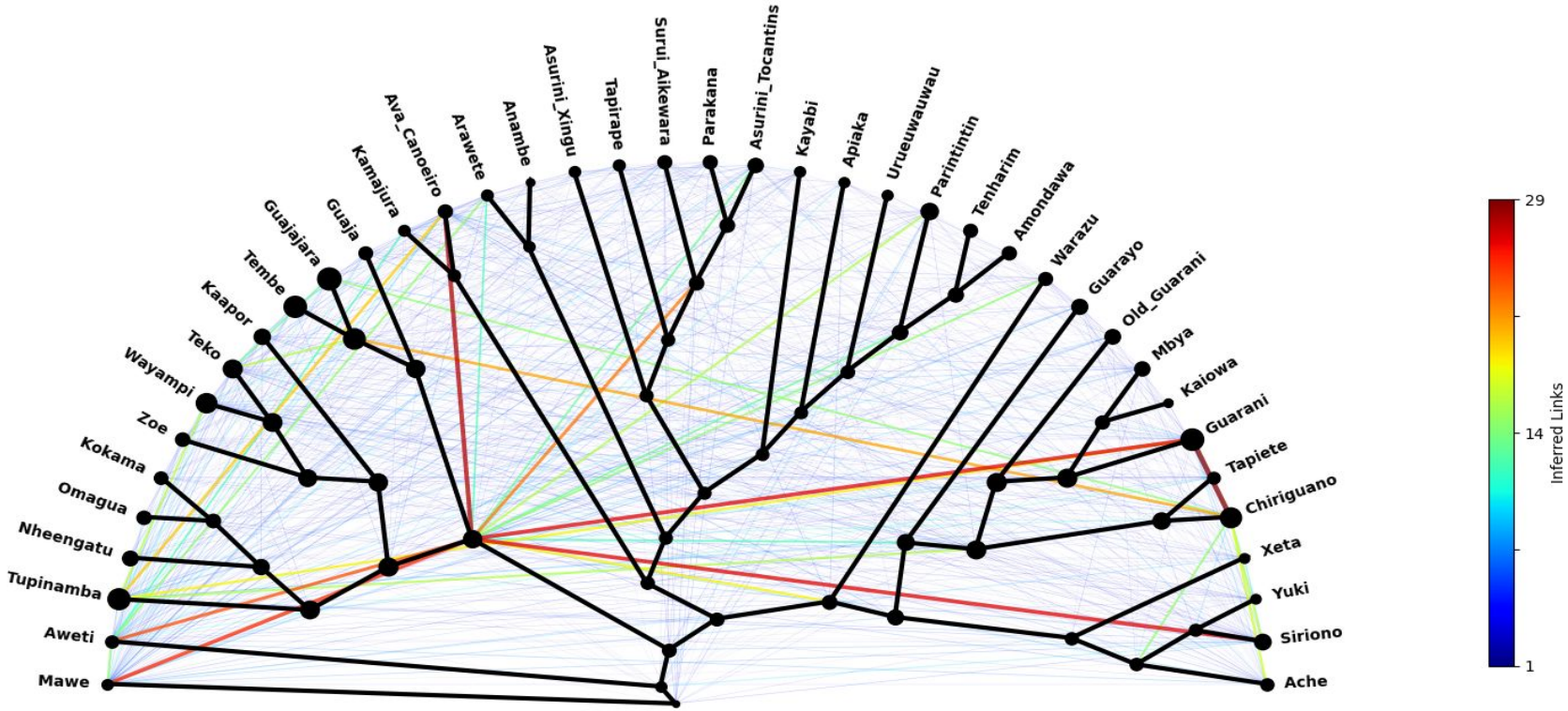
- Adds a new parameter for Bayesian inference
- While the BEAST2 package is ready for usage, the auxiliary scripts are still tailored for Indo-European



Methods - III

- We took a step back and revisited minimal lateral networks (MLNs)
- Starting to develop our own approach that uses the edges identified by MLNs
- Code under *intense* development

Minimal Lateral Network

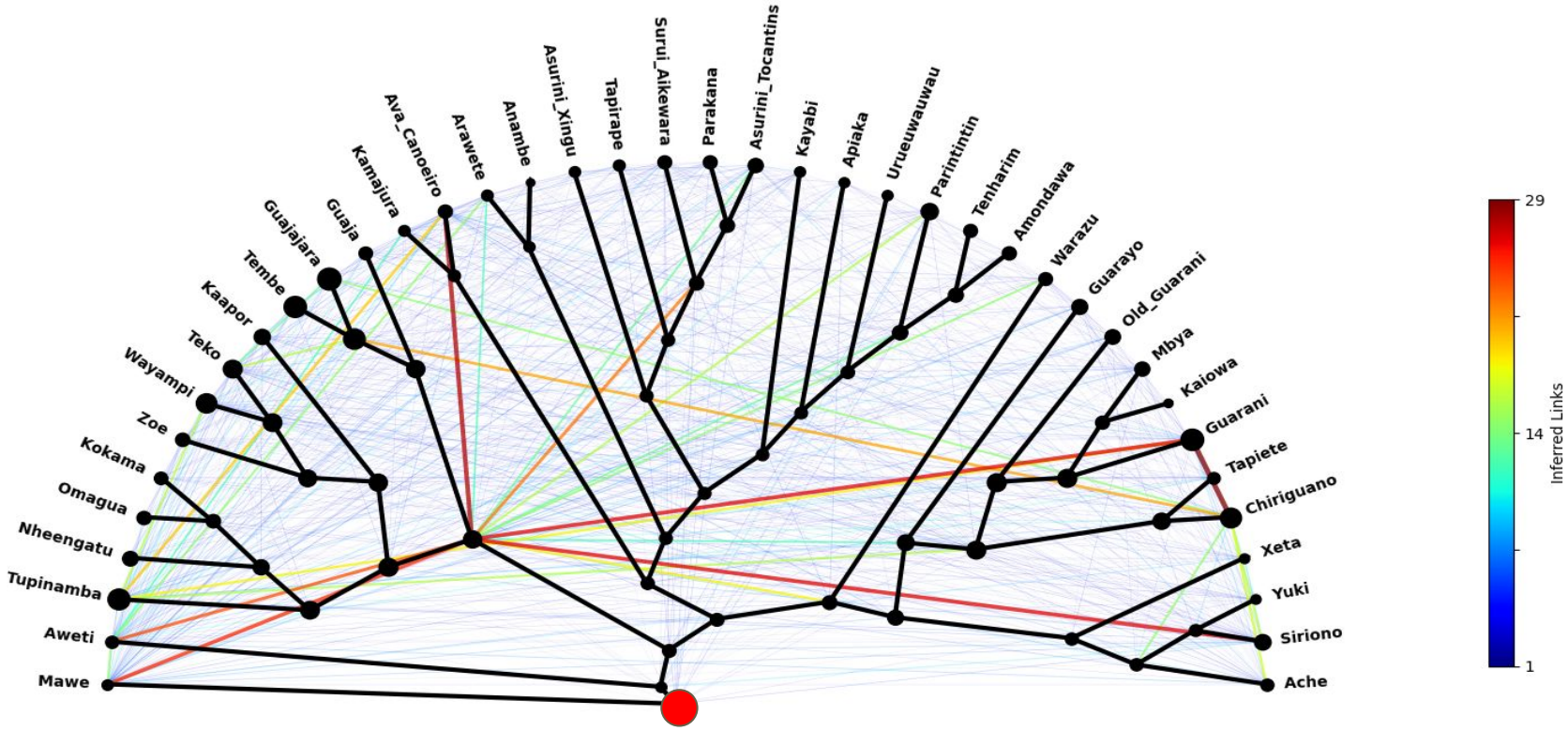


Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Stronger edges

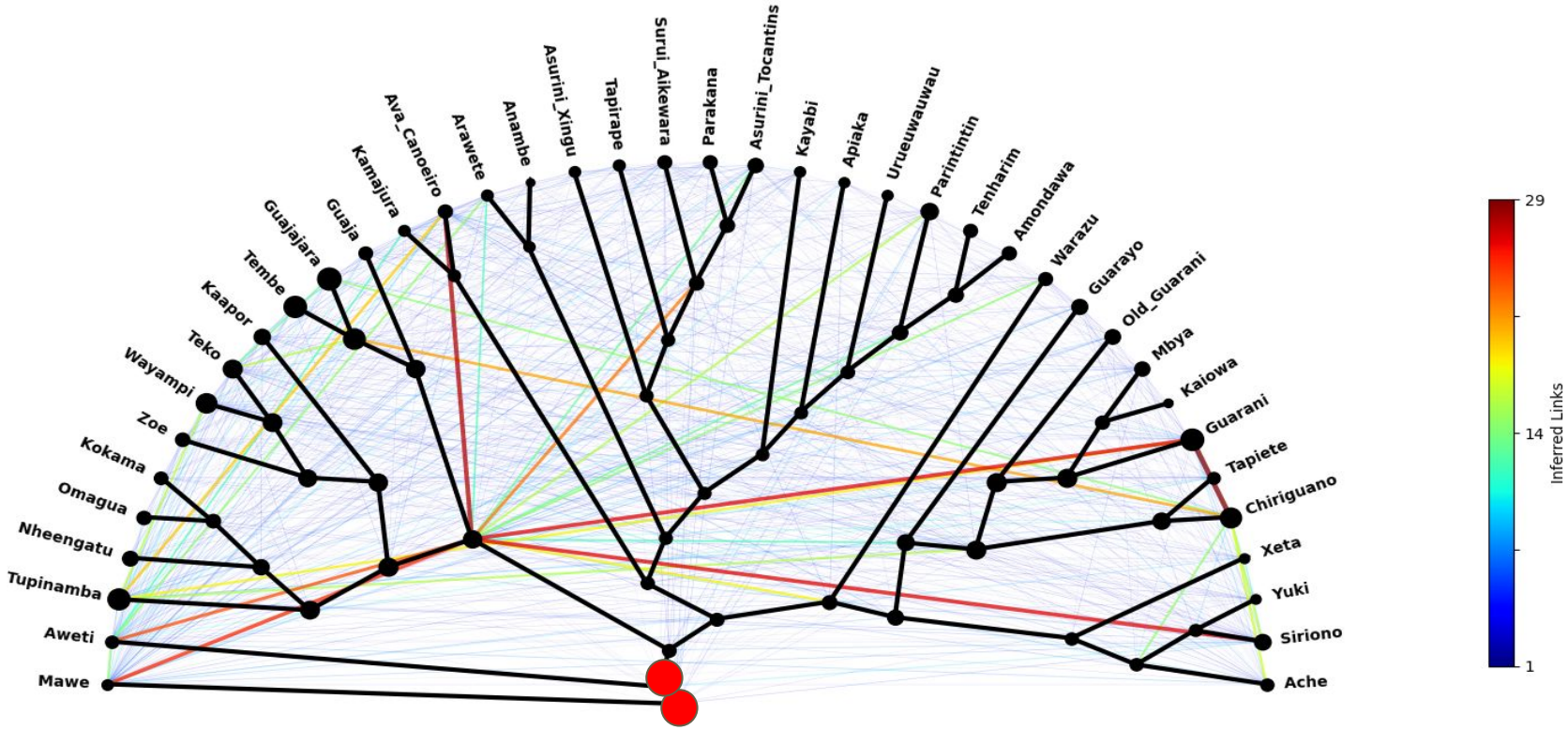
Group A	Group B
Chiriguano	Guaraní
Avá-Canoeiro	North and East group
Siriono	North and East group
Guaraní	North and East group
Mawé	North and East group
Aweti	North and East group
SPAt	North and East group
Chiriguano	Tenetehara (Guajá and Tembé)
Avá Canoeiro	Tupinambá

Minimal Lateral Network



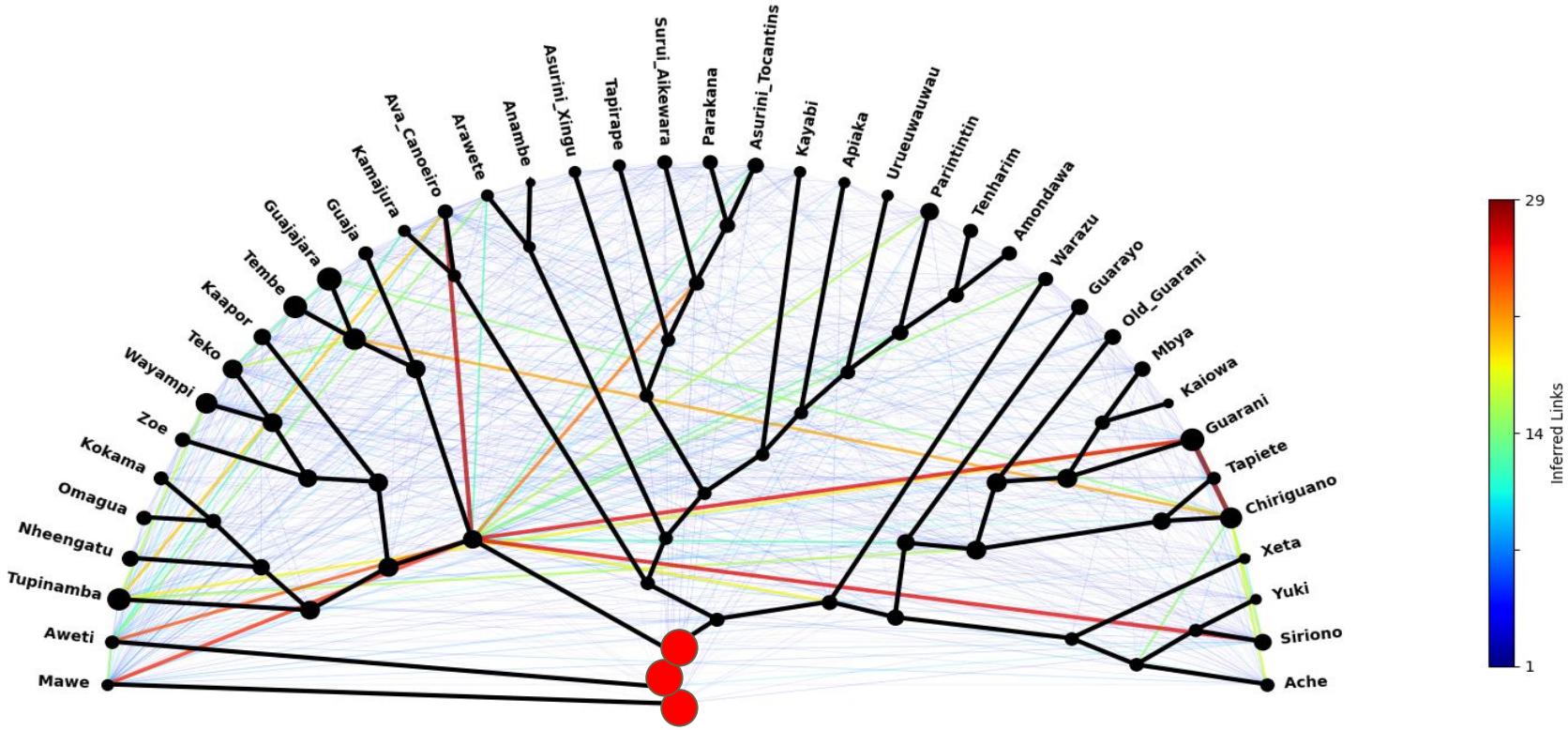
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



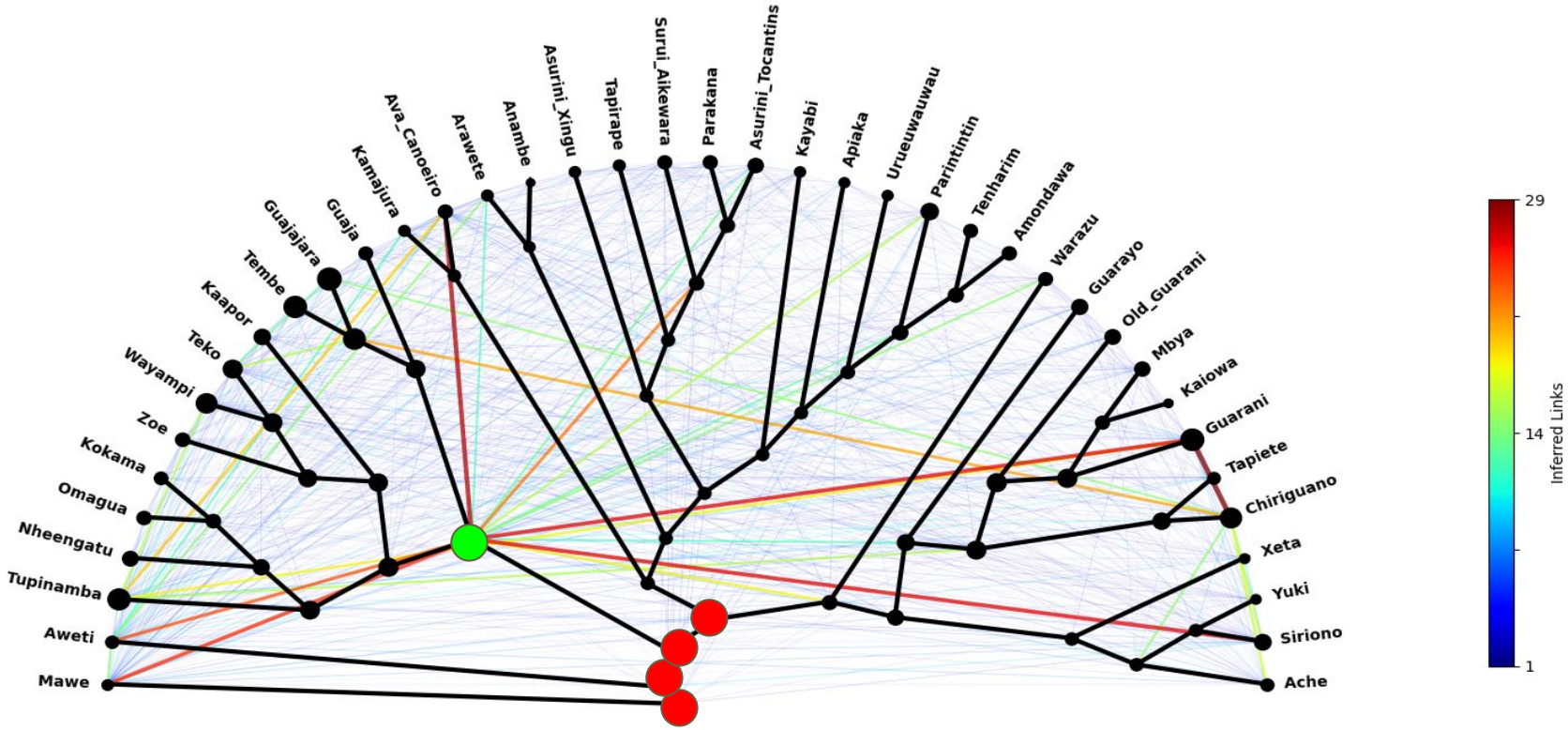
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



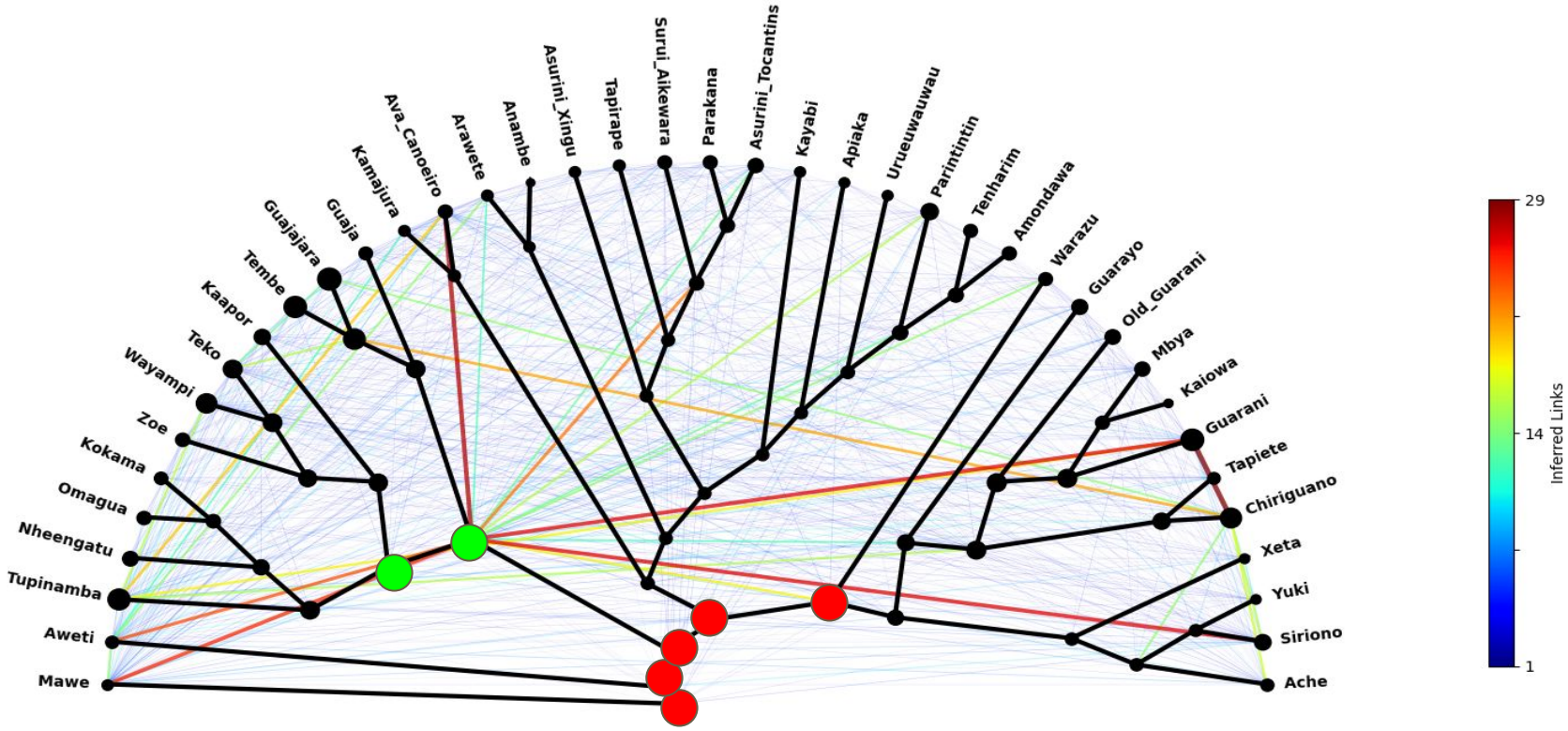
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



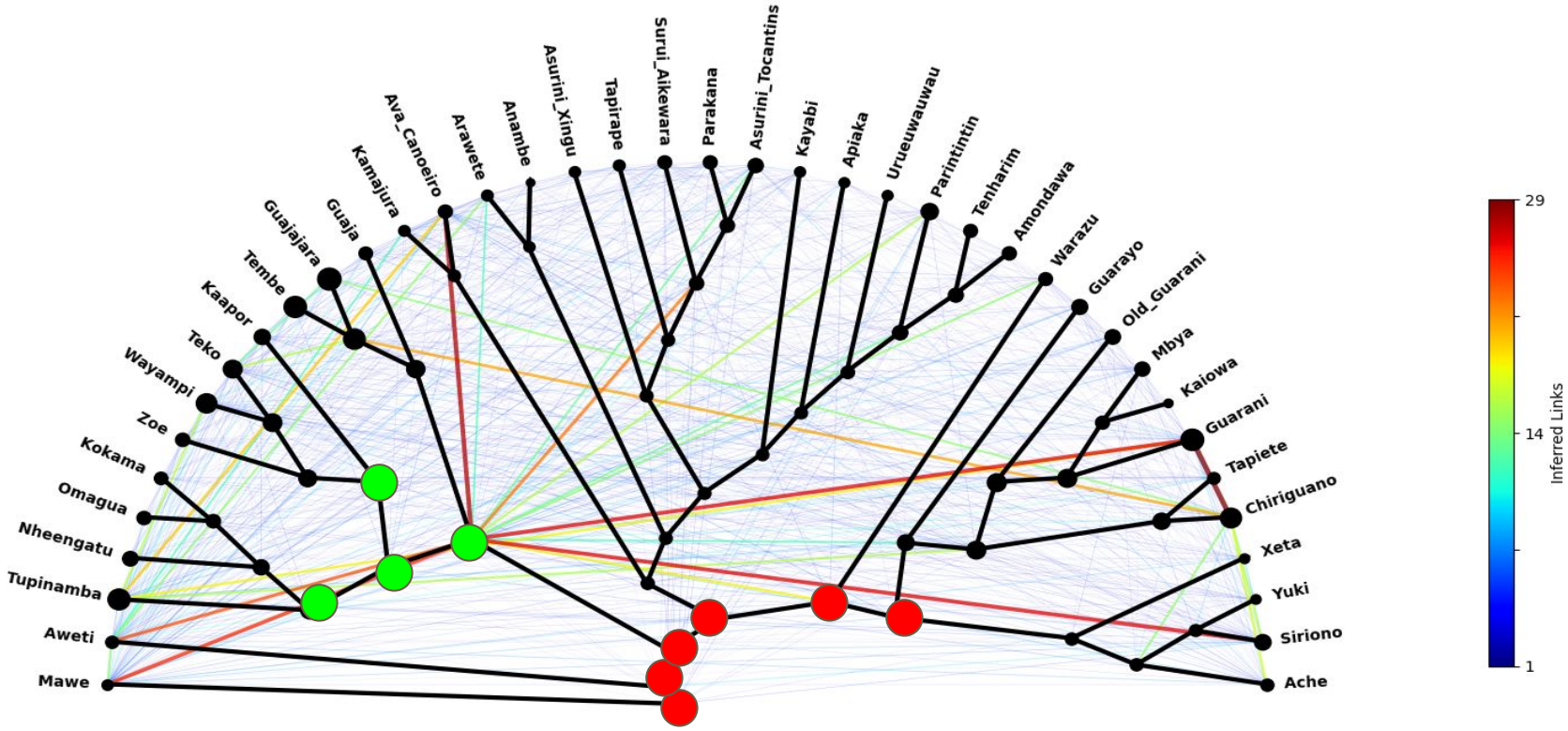
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



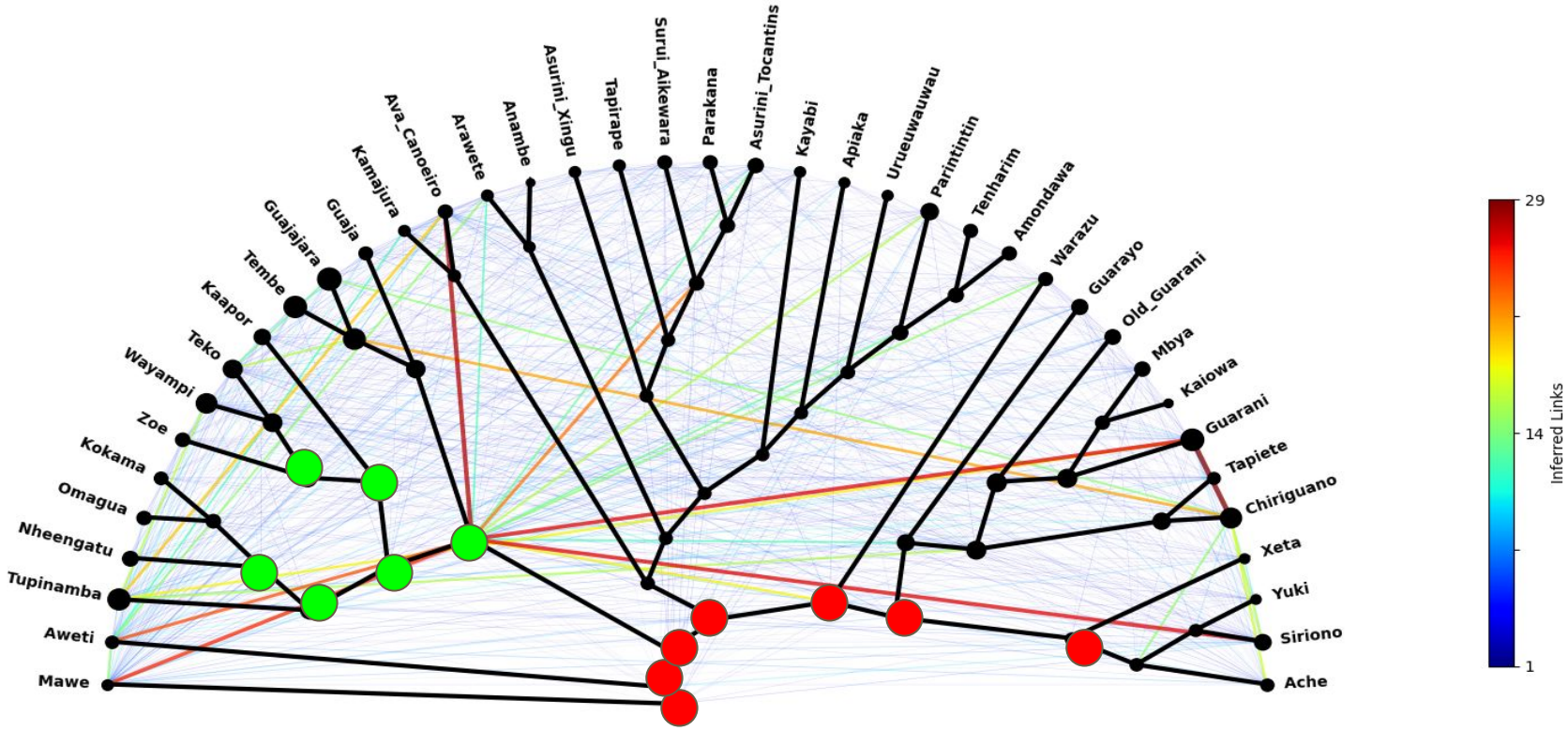
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



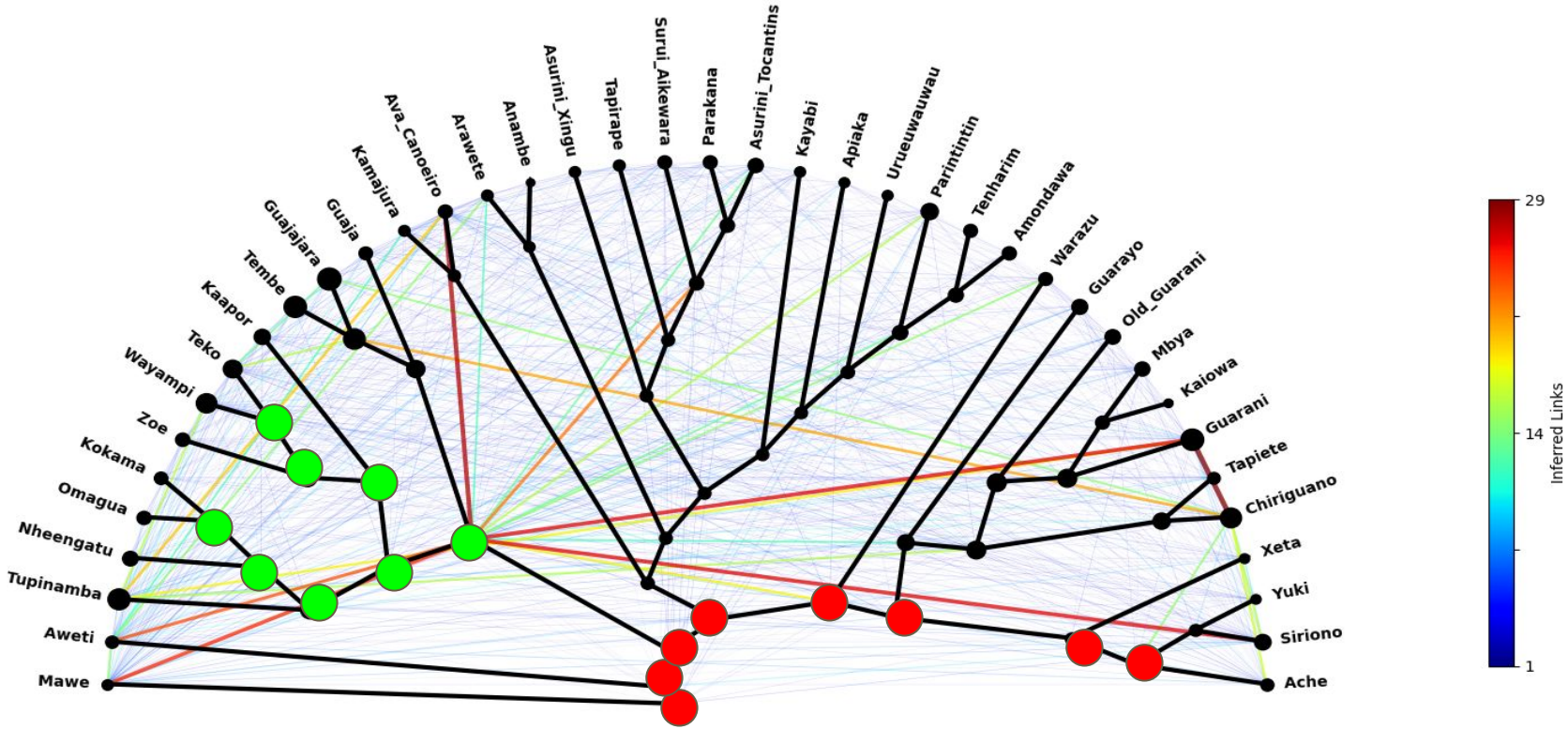
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



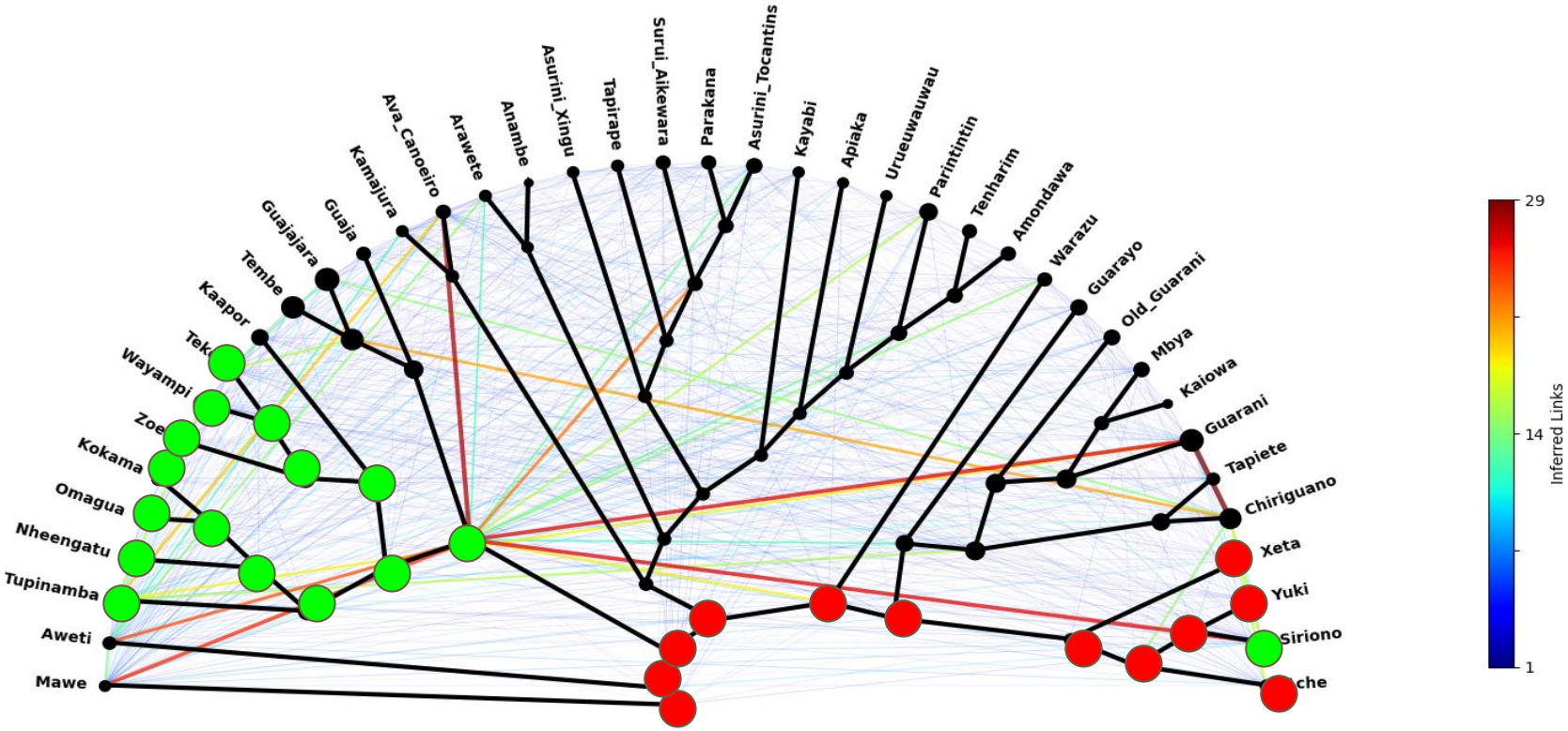
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



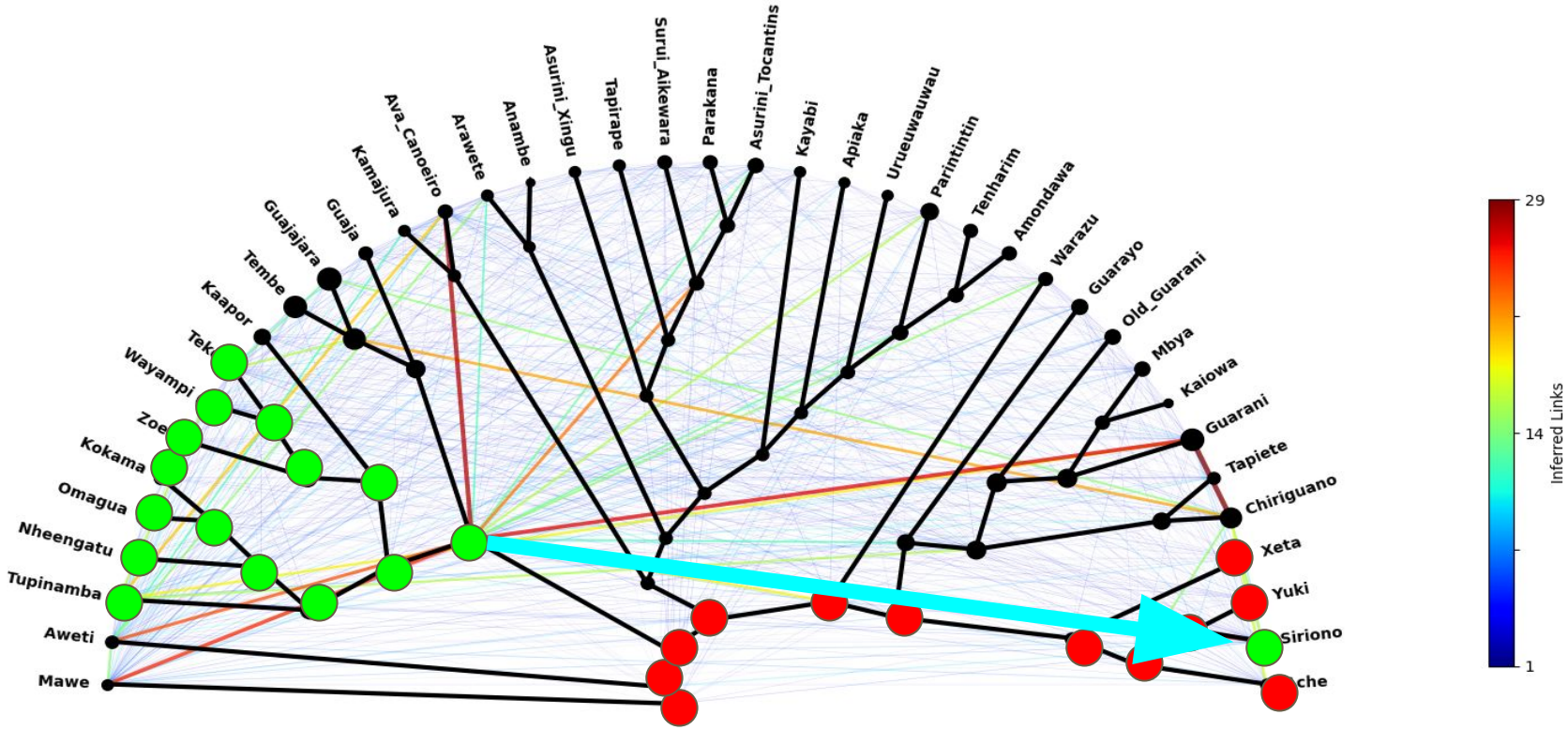
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



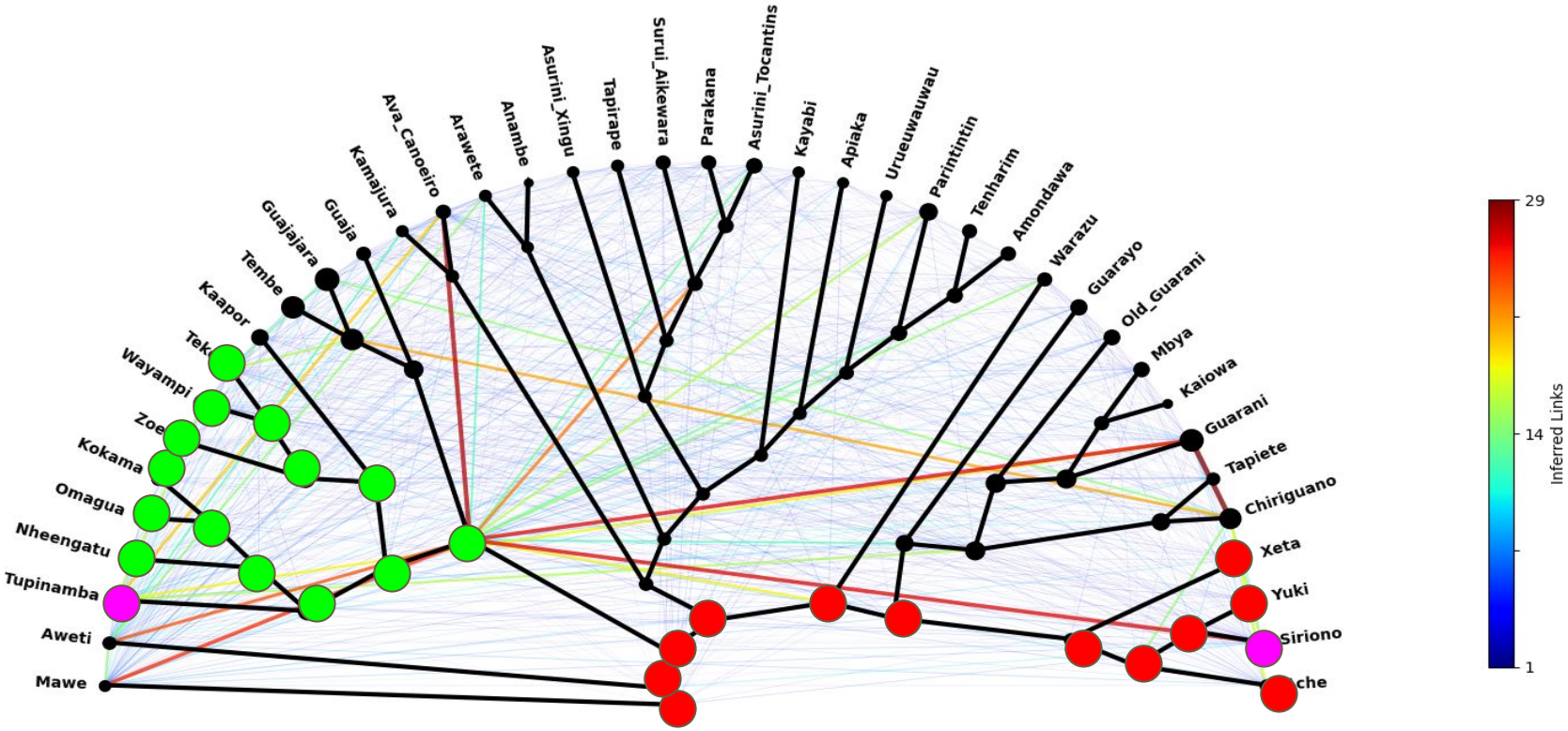
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



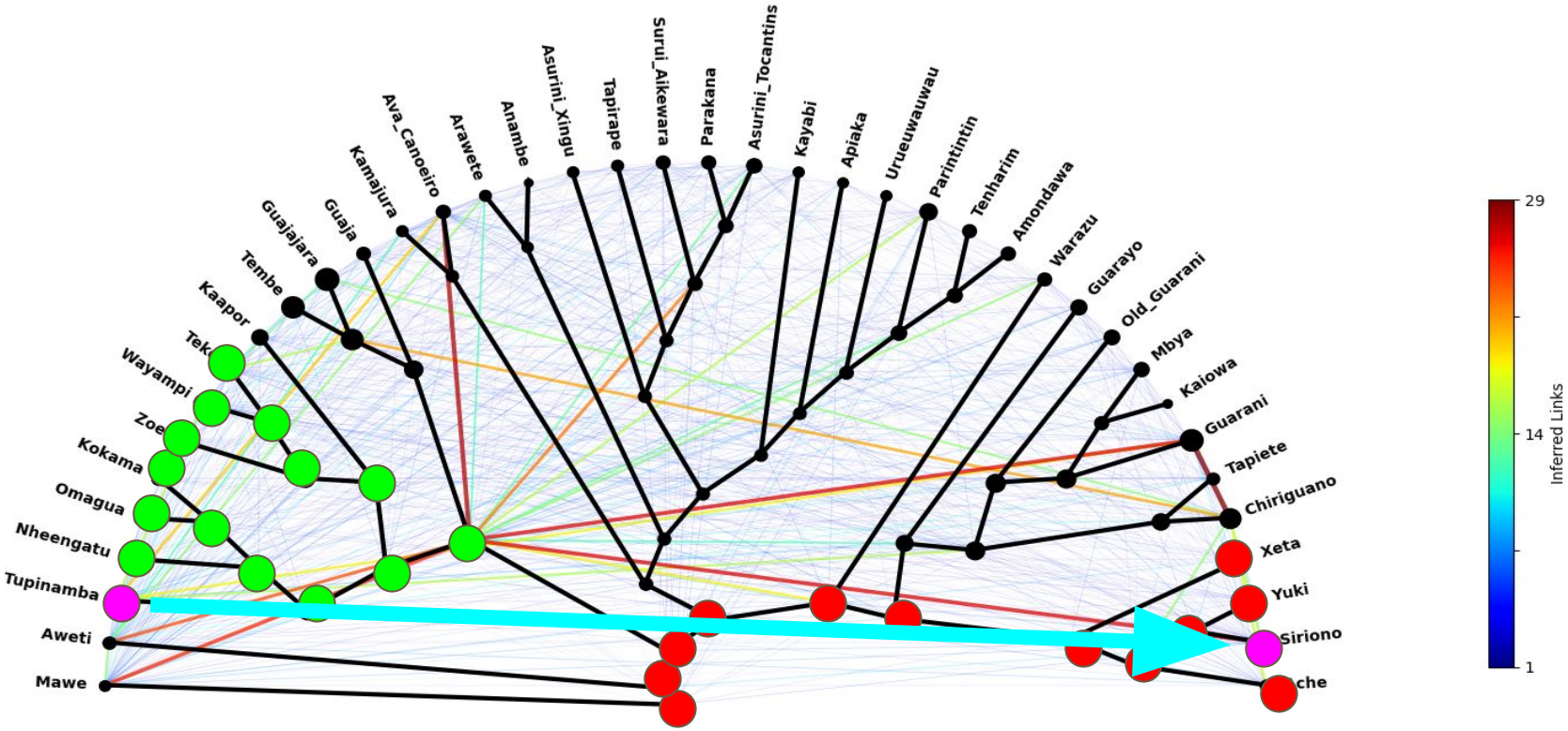
Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Minimal Lateral Network



Built with code adapted from Nelson-Sathi et al. (2011) and List et al. (2013).

Examples: Guaraní-Chiriguano

- /apɨ̃wa/ ‘nostril’
- /peu/ ‘pus’
- /tuku/ ‘grasshopper’
- /kʷatia/ ‘paper’
- /dʒuː/ ‘needle for sewing’

Examples: Northern and Eastern groups-Avá Canoeiro

- /kʷãnu/ ‘hedgehog’
- /wajnɨmɬ̃/ ‘hummingbird’
- /winga/ ‘inga (genus)’
- /uɲua/ ‘pestle’
- /poɨʁa/ ’bead’
- /amu/ ‘snow’
- /ɨŋ/ ‘shadow’
- /tĩ/ ‘plant something’

Examples: Northern and Eastern groups - Mawé

- /ʃ̃uʔi/ ‘hedgehog’
- /waipaka/ ‘hen’
- /mamãw/ ‘papaya’
- /amẽduĩ/ ’peanut’
- /pir+ãja/ ‘piranha’
- /jurumũ/ ‘pumpkin’
- /kuja/ ‘gourd’
- /arawe/ ‘cockroach’

Conclusion

- Results support the hypothesis of a partially reticulate evolution of Tupí-Guaraní
- Whether the network is an appropriate representation of the evolution of Tupí-Guaraní remains an open question
- Laying foundations for similar research with other families of the same geographic area

Thank you!


amilies
ana-Purubora
rukú
pían
uaraní
Uruku
Urumi
Ntogapid Itogapuk
Ramarama
Puruborá
Surui Paiter
Matanau
Kabanae
Monde
Cinta Larga
Gaviao Ikoloehj
Zoro
Arua
Aruashi
Kepkiriwat
Makurap
Karitiana
Arikem
Tupari
Wayoro
Akuntsu
Mekens Sakurabia
Kampe
Munduruku
Kuruaya
Xipaya
Juruna Yudja
Satere Mawe
Aweti
Tupi do Machado
Kamajura
Asurini do Xingu
Tapirape Apyawa
0.0°
10.0°S