

---

# A NETWORK ALIGNMENT BASED METHOD FOR PHYLOGENETIC ANALYSIS OF INTERDEPENDENT MORPHOLOGICAL CHARACTERS

---

April 8, 2021

Tristan J. Canterbury  
Supervisor: Dr. Martin D. Brazeau  
Imperial College London  
MSc Computational Methods in Ecology and Evolution  
`tjc19@ic.ac.uk`

# 1 Keywords

1. Phylogenetics
2. Morphology
3. Software development
4. Hierarchical characters
5. Dynamic homology
6. Comparing networks

# 2 Introduction

Phylogenetic inference based on morphology allows us to corroborate proposed phylogenies of extant lineages with evidence from the fossil record for which we only have morphological evidence. As most species that have ever lived are extinct, few of which we have fossils for, the value of robust methods for inferring phylogenetic relationships from morphology becomes apparent. In particular, problems that do rely heavily on morphological evidence, such as the long unresolved Arthropod head problem [Budd, 2002, Rempel, 1975], might benefit from fresh methods of phylogenetic inference. An issue that arises when we attempt to measure phylogenetic distances based on morphological similarities, from which homology is somewhat arbitrarily inferred, is that many morphological characters do not evolve independently and tend to form hierarchical structures [Hopkins and St. John, 2021]. Taking a wide glance at biology we should expect morphological characters to be interdependent for many different reasons: All morphological characters grow from the same embryo, they are physically connected, most are polygenetic, most alleles that code for a particular character will also code for another (pleiotropy), and they all co-evolve and they may serve similar adaptive purposes and so may converge on similar features. Whilst many of these relationships are hard to ascertain, the physical relationships between these characters are often obvious

but underutilised in current quantitative methods of phylogenetic analysis as there has been no good way to codify them.

The aim of my project is to represent morphological interdependencies as a network/graph of characters, an anatomical network, and infer phylogenetic distance based on evidence of transformation events between the topologies of these networks, with the assumption that fewer graph transformations means greater parsimony [Grant and Kluge, 2004]. Further questions to explore would be to ask what other useful information can be gained from anatomical networks besides phylogenetic distance, how effective is this method in comparison to other methods and how can the software be further optimised.

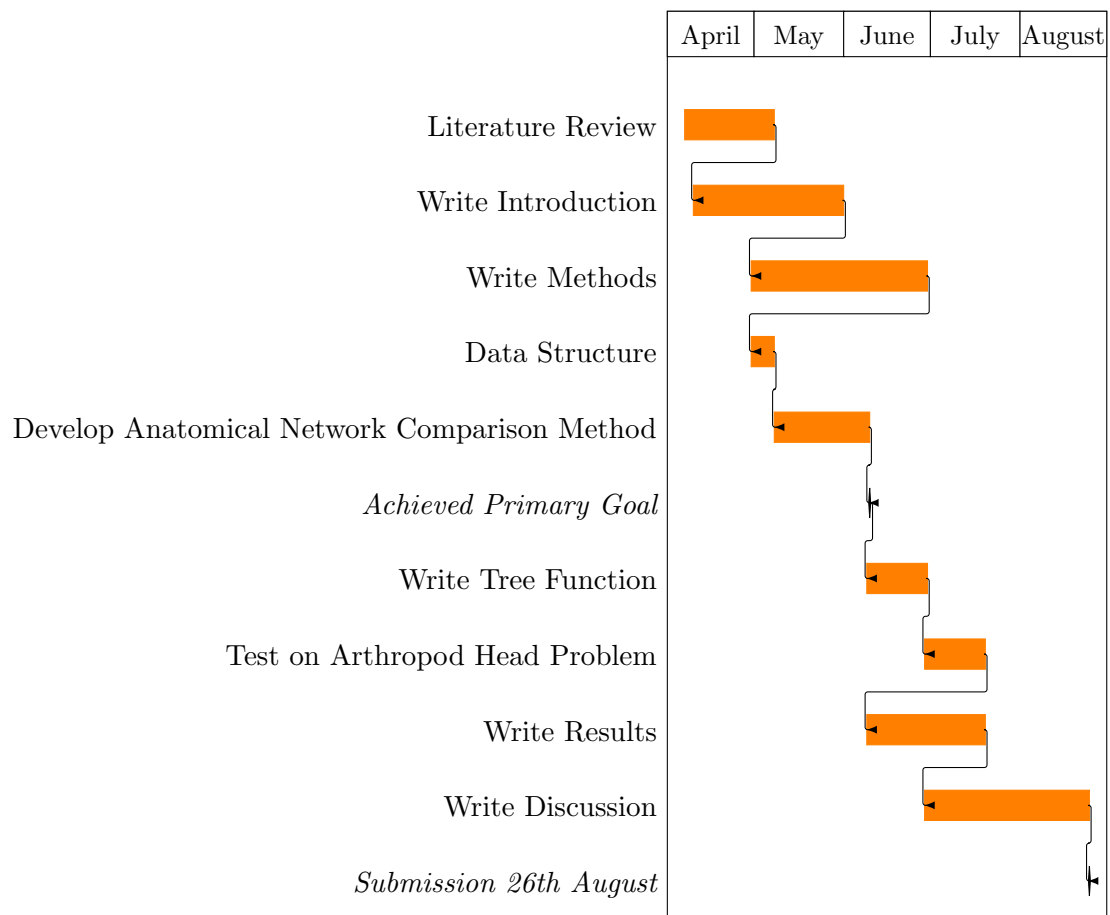
### **3 Proposed Methods**

With my goal being to reconstruct ancestral states without making prior assumptions of homology between particular characters, this method of comparing networks must work where there is an unknown node-correspondence (UNC). These node correspondences will instead be estimated using a alignment-based network comparison method. A measure of distance between networks is also needed so that I can infer the most likely phylogenetic relationships between nodes of the tree. Such methods generally belong to the GRAAL family as these algorithms estimate node correspondences based on the topology of the network and outputs a measure of distance between the trees [Tantardini et al., 2019]. Of the GRAAL family the Lagrangian graphlet-based network aligner (L-GRAAL) is currently the best candidate as it also allows for use of biological information and directional graphs [Malod-Dognin and Pržulj, 2015]. The main drawback of these methods is that their computational efficiency scales quadratically or worse with the number of nodes. I may have to construct my own prototype of this software in python for morphological data however as the C++ source code for the L-GRAAL programme is not available for me to adapt for these purposes.

## 4 Anticipated Outputs and Outcomes

- Development of a method for measuring phylogenetic distances between anatomical networks.
- A prototype software package in python and C++ for implementing this new method for tree evaluation.
- An analysis of the efficacy of this method for solving the problem of morphological character interdependencies through the example of the Arthropod head problem.

## 5 Project Feasibility



## 6 Budget

Item	Price (£)	Justification
4 TB External Hard Drive	100	Safe storage of project files.
Books	300	References and courses on set theory, graph theory, Information Theory, complex networks and phylogenetic analysis.

## 7 Supervisor Declaration

**I have seen and approved the proposal and the budget.**

Primary Supervisor: Dr Martin D. Brazeau

Signature:

Date: 08/04/2021

## References

- [Budd, 2002] Budd, G. E. (2002). A palaeontological solution to the arthropod head problem. *Nature*, 417(6886):271–275.
- [Grant and Kluge, 2004] Grant, T. and Kluge, A. G. (2004). Transformation series as an ideographic character concept. *Cladistics*, 20(1):23–31.
- [Hopkins and St. John, 2021] Hopkins, M. J. and St. John, K. (2021). Incorporating Hierarchical Characters into Phylogenetic Analysis. *Systematic Biology*, 0(0):1–18.
- [Malod-Dognin and Pržulj, 2015] Malod-Dognin, N. and Pržulj, N. (2015). L-GRAAL: Lagrangian graphlet-based network aligner. *Bioinformatics*, 31(13):2182–2189.
- [Rempel, 1975] Rempel, J. (1975). Evolution of the Insect Head: the Endless Dispute. *Quaestiones Entomologicae*, 11:7–25.

[Tantardini et al., 2019] Tantardini, M., Ieva, F., Tajoli, L., and Piccardi, C. (2019).  
Comparing methods for comparing networks. *Scientific Reports*, 9(1):1–19.