PajaroLoco: A suite of programs to study complex adaptive properties of animal language. An example of a Cassin's vireo syntax network.

Héctor Manuel Sánchez Castellanos[1] and Edgar E. Vallejo [1] and Charles E. Taylor [2]

[1]Computer Science Department,Tecnológico de Monterrey, Estado de México, 52926, México.
[2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California, 90095, USA.
sanchez.hmsc@itesm.mx

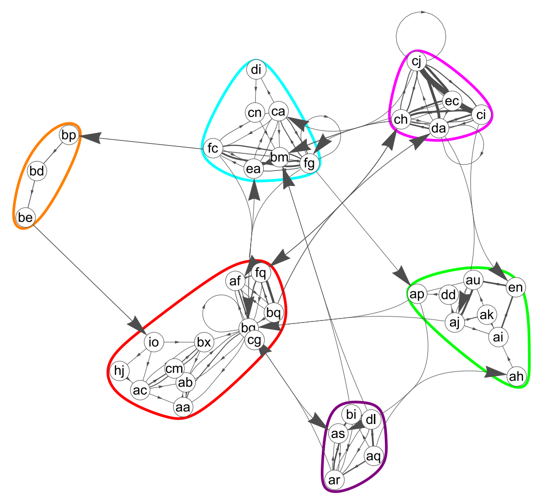
Introduction

Animal language may be regarded as a complex adaptive system. Although there is software for identifying linguistic units or "phrases" in animal vocalisations, packages for analysing the grammatical properties of phrases sequences are scarce. *PajaroLoco* is an open-source *Mathematica* package for the study of these features. This paper is a demonstration of the capabilites of *PajaroLoco* using as an example the syntax network from a Cassins Vireo individual. Our intention is to illustrate how it could be used for other network analysis tasks in the artificial life community. It should be noted, though, that it is not our purpose to perform a thorough description of the species grammar as we are currently working on other aspects of that field of study (Arriaga et al., 2015; Hedley, 2015).

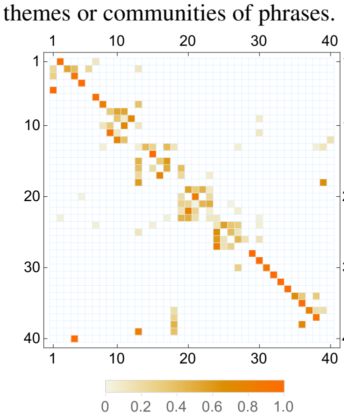
Bird song Analysis

We will describe one song from a Cassins Vireo male, "Mine" (sample id: 1156), singing 559 phrases on 13th of May of 2013 (for a description of such songs and a link to recordings see Hedley (2015)). An excerpt of the annotated sample is: *au, aj, ak, ai, ..., da, cj, ch, ci*; where each pair of letters represents a distinct sound composition (these phrases were obtained *a priori* by analysing its spectrogram with different software) and the position of the phrases in the sequence denotes temporal relationships (*au* occurs before *aj*, *ak* before *ai*, etcetera). All of the routines that will be shown can be performed directly in *PajaroLoco* (Sanchez et al., 2015).

Network Representation and Small-World Themes: The phrase sequences in the song were first represented as a network. This was achieved by setting the phrases as vertices in the graph and the transitions between them as weighted edges (an animation with sound of how this is done can be seen on the project's youtube playlist: `https://www.youtube.com/playlist?list=PLRzY6w7pvIWrQXIcnxN5tKtpIVXjhqOlu`). In figure 1a we can see the resulting network. It is easy to observe the tendencies of certain phrases to appear in communities or "themes". How tightly these themes are

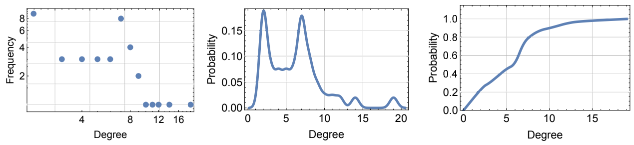


(a) Network representation of the transition frequencies of the song. Vertices are the phrases of the song and edges represent transitions. The coloured groups represent the themes or communities of phrases.



(b) Markov transitions probability matrix. The numbers on the frame are the identifiers of the phrases and the transition probability is represented by a color scale.

Figure 1: Markov and network representation of the song's transitions.



(a) Degree Frequencies. (b) Probability distribution. (c) Cumulative probability.

Figure 2: Vertices degrees plots.

connected can be measured with the "small-worldness" coefficient which in this case is: 3.23. A random graph would have a "small-worldness" of approximately 1 so the song of the bird is not structured in a completely random manner. Instead, phrases are organised into clusters in which a collection of phrases are often used together for a period of time. We also obtained the Markov transitions matrix (figure 1b). In it we can see how probable is that the bird vocalises a certain phrase after he has produced another given one. We can observe once again that phrases follow a pattern and are not sung at random. Hedley et al. (2016) have analysed these data further, and suggested that the best representation is someplace between a first- and second-order Markov process, though some transitions may be more complex.

Morphological Analysis: While most of the network analysis takes place in undirected networks some measures deal with the in and out-degree of the vertices. In this case they are grouped in patterns or "motifs" according to how they are connected to other phrases (bottlenecks, hourglasses, one-ways and branches (Sasahara et al., 2012; Cody et al., 2015)). In our example we used the mean number of edges (3.05) in the network as the threshold for the detection of structures (this is a rather arbitrary way to define the threshold for demonstration purposes but the parameter can be easily changed if desired). Hourglasses and one-ways were the most common with 9 phrases while bottlenecks and branches were less frequent with 6 phrases.

Degree Distribution: Another important feature to study in graphs is whether or not they conform to certain degree distribution patterns. Figure 2 shows three plots that were used for this purpose. We see in figure 2a that most phrases have few connections although we have another peak around seven connections hinting at a bi-modal distribution. This can be further observed in figure 2b. We can also see in figure 2c that most of the phrases (around 90%) fall between one and eight connections.

Conclusions and Future Work

Analysing animal language is important from a computational and complex adaptive systems point of view. Taylor and Cody remark this importance by emphasising the different varieties in which bird's vocalisations come and how they compare with other complex phenomenon such as cellular automata (Taylor and Cody, 2015). *PajaroLoco* is a tool developed for that purpose. Our program is part of an ongoing project and as such is updated and documented regularly. Although its main application is intended to be the analysis of annotated animal vocalisations, the package presented here can be used for the study of other complex adaptive phenomena in artificial life research, specially in those applications where phenomena can be described as sequences of elements and are amenable to network analysis.

Funding and Acknowledgments:

This work was supported by the US National Science Foundation under Award Number 1125423 and by Consejo Nacional de Ciencia y Tecnología under Award Number I010/214/2012. We thank Richard Hedley for permitting us to use his recordings and annotations of Cassins Vireo songs.

References

Arriaga, J. G., Sanchez, H., Vallejo, E. E., Hedley, R., and Taylor, C. E. (2015). Identification of Cassin's Vireo (Vireo cassinii) individuals from their acoustic sequences using an ensemble of learners. *Neurocomputing*, pages 1–14.

Cody, M. L., Stabler, E., Sánchez Castellanos, H. M., and Taylor, C. E. (2015). Structure, syntax and small-world organization in the complex songs of California Thrashers (Toxostoma redivivum). *Bioacoustics*, 4622(November):1–14.

Hedley, R. W. (2015). Composition and sequential organization of song repertoires in Cassin's Vireo (Vireo cassinii). *Journal of Ornithology*, 157(1):13–22.

Hedley, R. W. et al. (2016). Complexity, Predictability and Time Homogeneity of Syntax in the Songs of Cassin's Vireo (Vireo cassinii). *Plos One*, 11(4):e0150822.

Sanchez, H., Vallejo, E. E., and Taylor, C. E. (2015). PajaroLoco: A suite of programs to study the grammatical structure of bird songs. In *Proceedings of the Twentieth International Symposium on Artificial Life and Robotics*, volume 2015, pages 148–153.

Sasahara, K., Cody, M. L., Cohen, D., and Taylor, C. E. (2012). Structural design principles of complex bird songs: a network-based approach. *PloS one*, 7(9):e44436.

Taylor, C. E. and Cody, M. L. (2015). Bird song: A model complex adaptive system. In *Proceedings of the Twentieth International Symposium on Artificial Life and Robotics*, volume 2015, pages 2–7.