Symbolic co-occurrence networks in bird calls, human music, and other language forms

M.C.M. Jamerlan, G.A. Tapang

National Institute of Physics, University of the Philippines Diliman

We characterize and compare symbolic co-occurrence networks in animal sounds, language forms, and music networks. Finding common motifs and degree distribution patterns in these symbolic networks may provide insight to the evolutionary history of these forms. Music, for example, remains a problem to scientists since it provides no clear direct developmental advantage. A previous study has discussed a possible shared pre-cursor (a protolanguage) for music and language characterized by an intermediate evolutionary communication system more similar to music than known spoken language [1]. Previous comparative studies focus more on how humans perceive and process them as information [2].

A complex network is a system of interconnected components called nodes. These components are connected by edges, which may represent any form of relationship. Complex networks have been used on written language by treating texts as interconnected systems, and analyzing semantic and structural characteristics [3].

Bird calls were obtained from various sources, and transcribed musical scores of Filipino folk songs were obtained from the University of the Philippines. The nodes in the co-occurrence networks represent each note (pitch) in the bird call or folk song, and weighted edges denote the duration between consecutive notes.

Poems were acquired from PoemHunter.com, and prose from Project Gutenberg. Letter networks were made from texts not included in previous data sets. Chinese character (logograms) networks were constructed from a translated Bible online. The text was divided into 1,180 chapters to produce parts of comparable length. Unique words, letters, or characters were represented as nodes, and are connected by an edge if they appear consecutively in the text.

The mean degree distribution of letter, poetry, and Chinese character networks followed a power law. Thus, the most common elements in each text have exponentially higher frequencies than the rest. Texts with lesser unique components have exponentially higher word repetitions than larger texts.

Music networks have more repetitive small subgraphs compared to language networks, which confirms that more complex bird calls are characterized by repetitive notes [4], and folk songs by short, repetitive melodic patterns. Chinese characters, prose, and poetry networks have comparable average shortest path lengths. Thus, Chinese characters behave more like words and phrases (semantics) than letters (phonetics) [5].

[l] N. Masataka, Phys. Life Rev. **6**, 11 (2009).

[2] M. Besson, D. Schon, Ann. N.Y. Acad. Sci. **930**, 232 (2001).

[3] J. P. Monsanto et al., IJMPC **25**, 1 (2014).

[4] E. Vilches et al., ICPR '06 (2006).

[5] Z. Wu, G. Tseng, JASIST **44**, 532 (1993).