The community structure of word association graphs

András Bóta\textsuperscript{a}, László Kovács\textsuperscript{b}, Miklós Krész\textsuperscript{c}, András Pluhár\textsuperscript{d}

\textsuperscript{a}University of Szeged, Institute of Informatics
bandras@inf.u-szeged.hu
\textsuperscript{b}University of West Hungary, Department of Applied Linguistics
klaszlo@btk.nyme.hu
\textsuperscript{c}University of Szeged, Gyula Juhász Faculty of Education
kresz@jgypk.u-szeged.hu
\textsuperscript{d}University of Szeged, Institute of Informatics
pluhar@inf.u-szeged.hu

Abstract

One of the defining characteristics of small-world networks is that their edge distribution is globally and locally inhomogeneous: nodes form dense groups inside the networks and these groups are called communities [1]. While the traditional definition of communities only allows disjoint subsets of nodes, the detection of overlapping communities became a prominent research subject in recent years [2]. In this presentation we will use the hub-based overlapping community detection method of Bóta et al. [3] to examine and compare the community structure of two word association graphs based on two different languages: English and Hungarian. The English network was created from the University of South Florida word association norms [4]. The Hungarian network was constructed by László Kovács using data collected from internet users [5]. We will examine, around which specific words are communities formed – for example category names and words with general meanings. We will also examine if these specific words and their associated groups are similar in the English and Hungarian networks.

Keywords: network science, community, community detection, word associations

MSC: 05C82, 91F20

References