Group extraction for real-world networks: The case of communities, modules, and hubs and spokes

Lovro Šubelj, Neli Blagus and Marko Bajec

University of Ljubljana, Faculty of Computer and Information Science,
Tržaška 25, SI-1001 Ljubljana, Slovenia
[name.surname]@fri.uni-lj.si

Abstract. Large real-world networks are often composed of characteristic groups of nodes with common linking pattern. Densely connected groups known as communities are perhaps the most well studied example with diverse applications in various fields. However, most real-world networks also contain other characteristic groups of nodes, while some parts of networks reveal no significant groups at all. To overcome the latter, Zhao *et al.* have recently proposed a community extraction criteria that explicitly guards against extracting groups that are not statistically significant. We extend their framework to arbitrary groups of nodes and propose a local group quality criteria W. Let S be a group of nodes and let T represent its linking pattern. Criteria W favors links between S and T, and penalizes links between S and T^C, while disregarding links with both endpoints in S^C. For this talk we consider groups with $S = T$ that correspond to communities and groups with $S \cap T = \emptyset$ denoted modules, with hub and spokes as a special case (for $|T| = 1$). We apply the extraction framework to the famous American football social network and a software network compiled from JUNG network analysis library (among others). Expectedly, communities explain almost 70% of the structure in the case of the social network, while other groups account for less than 5% (according to the number of links $|L|$). On the other hand, communities contain only $\approx 12\%$ of the links in the software network, whereas modules, and hubs and spokes, account for $\approx 65\%$ of the structure. All extracted groups well coincide with the ground truth and the literature.

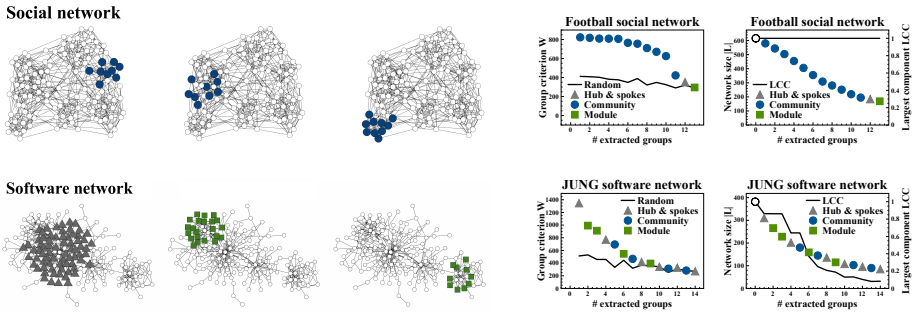

Fig . Group extraction from football social network and JUNG software network. Overlays show first three groups, while results are statistically significant at 1% level.

Keywords: real-world networks, community, module, hub and spokes, group extraction, social networks, software networks.

References

1. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. P. Natl. Acad. Sci. USA 99(12), 7821–7826 (2002)
2. Šubelj, L., Bajec, M.: Community structure of complex software systems: Analysis and applications. Physica A 390(16), 2968–2975 (2011)
3. Zhao, Y., Levina, E., Zhu, J.: Community extraction for social networks. P. Natl. Acad. Sci. USA 108(18), 7321–7326 (2011)