Accepted Manuscript

Validation of community robustness

Annamaria Carissimo, Luisa Cutillo, Italia De Feis

PII: S0167-9473(17)30234-7
DOI: https://doi.org/10.1016/j.csda.2017.10.006
Reference: COMSTA 6522

To appear in: *Computational Statistics and Data Analysis*

Received date : 21 March 2017
Revised date : 25 October 2017
Accepted date : 27 October 2017

Please cite this article as: Carissimo A., Cutillo L., Feis I.D., Validation of community robustness. *Computational Statistics and Data Analysis* (2017), https://doi.org/10.1016/j.csda.2017.10.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

*Final Revised Manuscript
Click here to view linked References

Validation of community robustness

Annamaria Carissimo

Bioinformatics Core, TIGEM, Pozzuoli, Italy

Luisa Cutillo[1]

University "Parthenope" of Naples, Italy, and University of Sheffield, United Kingdom

Italia De Feis

Istituto per le Applicazioni del Calcolo "Mauro Picone", Naples, Italy

Abstract

The large amount of work on community detection and its applications leaves unaddressed one important question: the statistical validation of the results. A methodology is presented that is able to clearly detect if the community structure found by some algorithms is statistically significant or is a result of chance, merely due to edge positions in the network. Given a community detection method and a network of interest, the proposal examines the stability of the partition recovered against random perturbations of the original graph structure. To address this issue, a perturbation strategy and a null model graph, which matches the original in some of its structural properties, but is otherwise a random graph, is specified. A set of procedures is built based on a special measure of clustering distance, namely Variation of Information, using tools set up for functional data analysis. The procedures determine whether the obtained clustering departs significantly from the null model. This strongly supports the robustness against perturbation of the algorithm used to identify the community structure. Results obtained with the proposed technique on simulated and real datasets are shown and discussed.

[1]Corresponding author: l.cutillo@sheffield.ac.uk, luisa.cutillo@parthenope.it

Download English Version:

https://daneshyari.com/en/article/6868839

Download Persian Version:

https://daneshyari.com/article/6868839

Daneshyari.com