Open Linked Social Data for Scientific Benchmarking

Renato Fabbri[a,1,*], Osvaldo Novais de Oliveira Junior[a,1]

[a] *São Carlos Institute of Physics, São Paulo University, Brazil*

Abstract

The field of social network analysis and the topic of complex networks are widely researched. Recently, a myriad of results have been reported which are based in diverse datasets most often not accessible to other researchers. This work exposes an open dataset with diverse provenance and oriented to provide the scientific community a friendly and common repertoire. Current data was obtained from Facebook, Twitter, IRC, Email and the specific instances of ParticipaBR, AA and Cidade Democrática. These were translated to linked data format to homonenize access, conform to current best practices and ease analyzes which integrate third party and provided instances. This document presents an outline and overall statistics of given dataset which should favor subsequent work.

Keywords: Benchmark Data, Facebook, Twitter, IRC, Email, Complex Networkx

1. Introduction

The enormity of the digital data makes a rapid development of various methods of data analysis, giving an opportunity to analyze the data from a different perspectives.

1.1. Benchmarking in the analysis of complexity

Karate club, whatelse?

*Corresponding author

Email addresses: fabbri@usp.br (Renato Fabbri), chu@ifsc.usp.br (Osvaldo Novais de Oliveira Junior)

[1] *URL:* http://www.ifsc.usp.br/

2. Materials: data from diverse provenance

3. Methods: linked open data

3.1. RDF

3.2. Data-driven ontology synthesis

4. Results

4.1. Data outline

4.2. Software tools

4.3. SPARQL queries

5. Conclusions

References