Ph.D. in "Life Course Research" – Socio-demographic curriculum

Academic Year 2023-2024

Statistical methods for complex data
27-30 May 2024
Maria Prosperina Vitale
mvitale@unisa.it

Objectives

The digital transformation and the diffusion of smart technologies allows scientists to collect data of increasing size and complexity. The rise of online social media is providing a wealth of different kind of textual, network and attribute data. The focus of this course will be on the statistical approaches to social media analytics for mapping, measuring and understanding the landscape of social media data in the modern world, discussing also the contribution of other disciplines ranging from social science to computer science in analyzing complex data structures derived from social media. More specifically, statistical tools will be provided to analyze complex data structures with an emphasis on: i) network measures and clustering on networks; ii) textual data analysis; iii) social media mining with supervised learning. Further aim is to present how complex data can be fruitfully exploited for the analysis of socio-demographic and educational phenomena. Case studies using various datasets will be provided. Moreover, students will gain hands-on experience utilizing the R software for implementing empirical analyses based on the methods discussed during the lectures.

Program

Monday, 27 May 2024

Monday afternoon (*Instructor*: *Vitale - Galluccio*) (3h) 14:00-17:00

- **Social Network Analysis**
Exploring network data structures
Clustering on networks

Tuesday, 28 May 2024

Tuesday morning (Instructor: Ragozini - Vitale) (3h) 10:00-13:00

-**The role of relationships in life course analysis**
Complex network data structures
Network models for peer effect and social influence in education

Wednesday, 29 May 2024

Wednesday morning (Instructor: Ragozini - Galluccio) (3 h) 10:00-13:00

-**Textual Data Analysis**
From texts to data

Correspondence analysis on lexical data
Sentiment Analysis and Topic modeling

Thursday, 30 May 2024

Thursday morning (Instructor: Lara Fontanella) (3 h) 9:30-12:30

- Social Media Mining and Supervised Learning for Textual Data
Social Media Mining: challenges and opportunities
Collecting data from social media platforms
Supervised Learning: Machine Learning vs Deep Learning Methods

Requirements
Elements of social network analysis, multilevel models, logistic regression models, multidimensional data analysis (Correspondence analysis and Cluster analysis)

Suggested lectures

Materials provided by the instructors

Further reading

Social Network analysis

Scott, J., & Carrington, P. J. (2011). The SAGE handbook of social network analysis. SAGE publications.

Genova, V. G., Giordano, G., Ragozini, G., Vitale, M. P. (2023). An analytic strategy for data processing of multimode networks. Advances in Data Analysis and Classification, 1-23.

Giordano, G., Ragozini, G., Vitale, M. P. (2019). Analyzing multiplex networks using factorial methods. *Social Networks*, *59*, 154-170.

Vitale, M. P., Porzio, G. C., Doreian, P. (2016). Examining the effect of social influence on student performance through network autocorrelation models. Journal of Applied Statistics, 43(1), 115-127.

Textual Data Analysis

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. " O'Reilly Media, Inc.".

Sharma, N. (2018, December). Sentiment Analysis using tidytext package in R. In *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)* (pp. 577-580). IEEE.

Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of statistical software*, *40*, 1-30.

Social Media Mining and Supervised Learning for Textual Data

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). Social media mining: an introduction. Cambridge University Press.

Hvitfeldt, E., Silge, J. (2022) Supervised Machine Learning for Text Analysis in R, Boca Raton, FL, CRC Press