

Title: Data mining in social network analysis

Author: Mgr. Peter Zvirinský

Department: Department of Theoretical Computer Science and Mathematical Logic

Supervisor: doc. RNDr. Iveta Mrázová CSc., Department of Theoretical Computer Science and Mathematical Logic

Abstract. In the past several years, the global economy has experienced a significant increase in overall debt, reaching 238% of the world GDP in 2022, as reported by the International Monetary Fund. This growing indebtedness raises concerns about the stability of the financial system and the welfare of individuals and institutions. It also underscores the need for effective strategies to understand the intricate relationships between debtors and creditors and to mitigate associated risks. In response, this thesis proposes a novel approach based on data mining methods for the comprehensive analysis of debt formation patterns among individuals and companies, focusing on the largely untapped data from the Insolvency Register (IR) of the Czech Republic.

We aim to leverage social network analysis (SNA) methods to model and analyze the interactions among subjects participating in insolvencies, namely debtors, creditors, and insolvency administrators. Additionally, we focus our research on dynamic social networks that capture structural changes in the data over time. Our approach enables an in-depth exploration of interactions, offering insights into debtors' and creditors' behavior and facilitating the prediction of future developments related to bankruptcies.

The methodology proposed in this thesis contributes to a better understanding of economic systems, identification of financial distress patterns, and facilitates decision-making in insolvency-related matters. Furthermore, the generic nature of the approach allows for its application to data from other insolvency registers and publicly available datasets, assuming similar preprocessing steps are undertaken.

Keywords: data mining, knowledge discovery, social network analysis, insolvency register, structured data, unstructured data