Our work deals with analysis of specific type of social networks called company networks. Actors are companies and connections are representing interactions between them. Dynamics of company networks is slower and they do not tend to create connections above certain boundaries (local business) because higher trust is needed than in other social networks. They also contain a lot of non-structural data as compositional data and temporal data. Company networks are interested already several projects in Slovakia like Foat.sk and ITUS.eu projects. These projects focused on visualization of ties in the network of companies in Slovak republic. These projects also partially utilize temporal data. In our work we build on the ITUS.eu project and enhance it in a number of ways. We proposed a comprehensive data model of company networks taking into account also aging factor (based on detailed temporal data). Main part of our work is dedicated to meso-level structural analysis, mainly analysis of local structures and local structure patterns detection, where approaches to large-scale network decomposition and positional analysis of local structures are designed and implemented.

Company Networks Analysis
Meso-level structural analysis

Martin Repka
Ján Paralič

Martin Repka
Martin Repka is a researcher focusing on new approaches and methods for specialisation social networking and analysis. He pays attention mainly to analysis and exploration of company networks. In his analyses he examines utilisation of data mining and knowledge discovery techniques in revealing and retrieval of new useful information in context.
Impressum / Imprint

Bibliographic information published by the Deutsche Nationalbibliothek: The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de. Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Coverbild / Cover image: www.ingimage.com

Verlag / Publisher:
LAP LAMBERT Academic Publishing
ist ein Imprint der / is a trademark of
AV Akademikerverlag GmbH & Co. KG
Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Deutschland / Germany
Email: info@lap-publishing.com

Herstellung: siehe letzte Seite /
Printed at: see last page
ISBN: 978-3-659-44207-0

Zugl. / Approved by: Košice, Slovakia, Technical University in Košice, Diss., 2013

Copyright © 2013 AV Akademikerverlag GmbH & Co. KG
Alle Rechte vorbehalten. / All rights reserved. Saarbrücken 2013
COMPANY NETWORKS ANALYSIS

Martin Repka, Ján Paralič

Abstract

This book is dealing with some approaches to analysis of specific social networks called company networks. These company networks emerge from interactions between companies, where a lot of these interactions are recorded by publicly accessible sources. Our work is designing and implementing methods of heterogeneous sources’ cleansing, aggregation and integration into comprehensive data model of company network. Later data transformations for analyses as various projections of company networks are designed, implemented and evaluated. Main part of our work is dedicated to meso-level structural analysis, mainly analysis of local structures and local structure pattern detection, where approaches to large-scale network decomposition and positional analysis of local structures are designed and implemented. All of the designed and implemented methods are experimentally tested on concrete instance of company network – Slovak Company Network.
Acknowledgement

This work was supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the grant No. 1/1147/12 and by the Slovak Research and Development Agency under the contract No. APVV-0208-10.
# Table of Contents

List of Symbols and Abbreviations.............................................................................6

1 Introduction ..............................................................................................................7
  1.1 Research goals ........................................................................................................8
  1.2 Book structure ........................................................................................................11

2 Theory of Social Networks.......................................................................................12
  2.1 Entities and Basic Concepts....................................................................................13
    2.1.1 Actor and Social Tie .......................................................................................13
    2.1.2 Dyad, Triad, Subgroup and Group .................................................................13
    2.1.3 Relation and Network ....................................................................................14
    2.1.4 Simple, Directed and Weighted Graph ............................................................14
    2.1.5 Significant Subgraphs and Cohesive Subgroups .............................................16
    2.1.6 Network Statistics .........................................................................................19
  2.2 Data for Social Networks ......................................................................................21
    2.2.1 Data Retrieval and Measurement ....................................................................21
    2.2.2 Network Data Representation .........................................................................22
    2.2.3 Structural, Compositional and Temporal data ...............................................23
  2.3 Types of Social Networks ......................................................................................25
    2.3.1 Types Based on Tie Characteristics .................................................................25
    2.3.2 Types Based on Network’s Actors ....................................................................25
  2.4 Real Network Models ............................................................................................27
    2.4.1 Small World Networks ....................................................................................28
    2.4.2 Scale Free Networks .......................................................................................29

3 Social Network Analysis .........................................................................................30
  3.1 Granularity of Analysis ..........................................................................................30
    3.1.1 Micro Level ......................................................................................................30
    3.1.2 Meso Level ......................................................................................................31
    3.1.3 Macro Level ......................................................................................................31
  3.2 Projections of Two-mode Networks ......................................................................32
    3.2.1 Projection of Affiliation Networks ....................................................................32
  3.3 Positional Analysis ................................................................................................35
    3.3.1 Roles ................................................................................................................35
    3.3.2 Equivalency ......................................................................................................37
5.1.2 Actors & Ties in SCRN ................................................................. 81
5.1.3 Comprehensive Model in SCRN .................................................. 82
5.2 Experiments ................................................................................. 84
5.3 Local Structure Analysis ................................................................. 85
  5.3.1 Subnetwork C2C LP* ................................................................. 85
  5.3.2 Initial Analysis of C2C LP* ........................................................ 87
  5.3.3 Getting Local Structure of C2C LP* .......................................... 87
  5.3.4 Analysing Common Local Structures ........................................ 90
  5.3.5 Blockmodeling on Local Structure ............................................ 93
  5.3.6 Conclusion ............................................................................ 94
5.4 Projection of Slovak Company Network ............................................. 95
  5.4.1 Experimental Projection PR1 ...................................................... 95
  5.4.2 Experimental Projection PR2 ...................................................... 99
  5.4.3 Properties of Projected SCCN .................................................. 100
  5.4.4 Conclusion ............................................................................ 103
5.5 Structure Pattern Mining ................................................................. 104
  5.5.1 LSPM Manager ....................................................................... 105
  5.5.2 LSPM Tryout .......................................................................... 106
  5.5.3 LSPM Experiments .................................................................. 108
  5.5.4 Interpretation of Resulting Patterns ........................................... 111
  5.5.5 Decomposition Strategy ............................................................ 113
  5.5.6 Conclusion ............................................................................ 114
5.6 Dynamics of Slovak Company Network .............................................. 114
  5.6.1 Snapshots of Projected SCN .................................................... 114
  5.6.2 Dynamics of Connections in SCN ............................................. 115
  5.6.3 Conclusion ............................................................................ 118
6 Goals evaluation and conclusion ............................................................. 119
  6.1 Final Remarks and Future work .................................................... 121
Bibliography ......................................................................................... 122
Appendices .......................................................................................... 128
Appendix A Additional descriptions of SCN ............................................ 129
  A.I Data of SCRN .......................................................................... 129
  A.II Network Statistics of C2C from SCRN and SCCN (PR2) ............... 132
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-MN</td>
<td>One-mode Network</td>
<td>One-mode Network</td>
</tr>
<tr>
<td>2-MN</td>
<td>Two-mode Network</td>
<td>Two-mode Network</td>
</tr>
<tr>
<td>C2C</td>
<td>Company to Company Network</td>
<td>Company to Company Network</td>
</tr>
<tr>
<td>CN</td>
<td>Company Network</td>
<td>Company Network</td>
</tr>
<tr>
<td>DSNA</td>
<td>Dynamic Social Network Analysis</td>
<td>Dynamic Social Network Analysis</td>
</tr>
<tr>
<td>FOAF</td>
<td>Friend of a Friend, project and also ontology</td>
<td>Friend of a Friend, project and also ontology</td>
</tr>
<tr>
<td>ICO</td>
<td>Identifikačné Číslo Organizácie (Official ID number of Company in SR)</td>
<td>Identifikačné Číslo Organizácie (Official ID number of Company in SR)</td>
</tr>
<tr>
<td>ITLIS</td>
<td>Information Transparency and Law Investigation System, project</td>
<td>Information Transparency and Law Investigation System, project</td>
</tr>
<tr>
<td>LSA</td>
<td>Local Structure Analysis, method and also experiment</td>
<td>Local Structure Analysis, method and also experiment</td>
</tr>
<tr>
<td>LSPM</td>
<td>Local Structure Pattern Mining, original method</td>
<td>Local Structure Pattern Mining, original method</td>
</tr>
<tr>
<td>ORSR</td>
<td>Obchodný Register Slovenskej Republiky (Business Register of Slovak Republic)</td>
<td>Obchodný Register Slovenskej Republiky (Business Register of Slovak Republic)</td>
</tr>
<tr>
<td>SCCN</td>
<td>Slovak Company Collaboration Network</td>
<td>Slovak Company Collaboration Network</td>
</tr>
<tr>
<td>SCN</td>
<td>Slovak Company Network</td>
<td>Slovak Company Network</td>
</tr>
<tr>
<td>SCPN</td>
<td>Slovak Company Projected Network</td>
<td>Slovak Company Projected Network</td>
</tr>
<tr>
<td>SCRN</td>
<td>Slovak Company Raw (Data) Network</td>
<td>Slovak Company Raw (Data) Network</td>
</tr>
<tr>
<td>SFN</td>
<td>Scale Free Network</td>
<td>Scale Free Network</td>
</tr>
<tr>
<td>SN</td>
<td>Social Network</td>
<td>Social Network</td>
</tr>
<tr>
<td>SNA</td>
<td>Social Network Analysis</td>
<td>Social Network Analysis</td>
</tr>
<tr>
<td>SWN</td>
<td>Small World Network</td>
<td>Small World Network</td>
</tr>
<tr>
<td>WCC</td>
<td>Weakly Connected Component, Weak (-ly connected) Component Clustering</td>
<td>Weakly Connected Component, Weak (-ly connected) Component Clustering</td>
</tr>
</tbody>
</table>
1 Introduction

Identification of purpose of connections between people is complicated and very difficult task. Social network analysis (SNA) deals also with this issue. It is even more complicated if we study specific connections like collaboration between companies. In business environment it is very important to have overview about these connections. All organizations have their statutory representatives, owners, and partner or stock holders. These are people who are involved in business and they can be affiliated with organizations e.g. companies. This view gives us idea to look at company as special social network, affiliate network actually.

As we see nowadays, all network technologies are claiming increasing significance in science community. The most evolving and spreading branch of them is mainly social networks and its analysis. Social network analysis (SNA) has become very common also amongst non-science oriented population as a part of their social life. This phenomenon increased a big interest in SNA. If we look at the Company Networks, we can find a lot of analogy with the social networks, or we can say that they are a specific kind of social network. Company Networks include Ownership, Partnership Networks and Collaboration networks, in which actors of the network are defined as well as relationships between them. On the web we can find a lot of sources containing data for Company networks (such as ORSR, FOAF, AZET and others), but the output is raw and hardly consistent. Amongst businessmen the knowledge that Company network can provide is very important. Very interesting in CN is structural analysis, such as revealing main paths (as a back bone of the network) or community detection (collaboration between organizations, sharing or not sharing the same resources) and clustering. With use of these kinds of analysis we can track down structure, which is beyond this network.

Unfortunately analysis of complex and unknown relationships is very difficult and usually can be done only the off-line way. Data are aggregated in some time interval and they are analysed using various methods, e.g. network graph or incidence matrix based methods. Sometimes these methods remove all dynamic attributes of connections and there are only dominant or obvious features visible. Despite this in SNA there are usually all time attributes
removed and in data temporal dimensions are removed too. This can be prevented using several approaches. First of all is to use method of dynamic social networks analysis (DSNA). In this case dynamic attributes are included as chronological order a there is possibility to solve process detection problem. Other option is to project dynamics of actors and connections of network in process of data transformation (e.g. projection) for SNA, where the temporal dimension is removed. This approach enables to utilize certain aspects of dynamics in classical SNA a support of new relationship discovery or presence of certain data correlation. The last simplest way is to transform temporal dimension in sort of attributes (weights) of actors and connections in process of data transformation to conform SNA.

Due to nature and significance of social networking data we can divide this data into several types, in this book we will deal with common structural data for network analysis and then we will incorporate other types of data to produce comprehensive data model of social networks. Representation of a comprehensive data model for social networks and different methods of composition analysis will be designed where fundamental aspects of positional analysis will be considered. Work will continue with the analysis of data sources for comprehensive data model. It analyses social networks of companies and evaluates some aspects of this specific social network.

1.1 Research goals

First aim of our work was to design and implement data model for company networks. This task includes data retrieval, data analysis, data integration into a common social network data model, creation of social network structure and all preparation for social network analyses. It is necessary to aggregate several sources of Slovak company network data, e.g.: Business Register on Internet (ORSR\(^1\)) and data from ITLIS\(^2\) project, various online company catalogues (Zlaté stránky\(^3\), Azet\(^4\) ...), and financial marker database Amadeus\(^5\). These sources contain multiple types of data which will be further incorporated into comprehensive data model for social networks. As a result of this goal, a formal

---

\(^1\) http://www.orsr.sk/
\(^2\) http://itlis.eu/
\(^3\) http://www.zlatestranky.sk/
\(^4\) http://www.azet.sk/
\(^5\) http://www.bvdinfo.com/Products/Company-Information/International/AMADEUS.aspx
model and suitable representation of this comprehensive model will be designed and implemented. JAVA language was used and as initial standard mark-up GraphML⁶ implementation and Java Universal Network / Graph Framework⁷ (JUNG) have been utilised.

**GOAL 1:** Design and implement comprehensive data model of company network.

After reaching comprehensive data model goal we can initialize and build company network. Company network will be mixed mode (combination of one mode and two mode network) and it will contain a lot of temporal data (as part of non-structural data). To process temporal data we need to introduce time function – aging (weighting) of actors and connections in company network. After considering actors’ duality in company networks we will need to design suitable methods for projection of affiliate mixed-mode company network on one-mode company network. This is necessary for subsequent analyses of companies within the network. We will assume that companies are more important for many types of analysis than any other type of actors in the network. As the result of this aim the method/process/algorithm necessary for transformation, integration of data and building, projection of company network is implemented.

**GOAL 2:** Aggregate, transform and utilize accessible data to create projected company network, evaluate contribution of such a projection.

As long as the company network will be created (success of **GOAL 1**) we can implement visualisation of company network, which could also show the benefits of comprehensive data model and data transformation to incorporate non-structural data. After visualization we will be able to examine structure of company network. We expect that company network should have outstanding structure, i.e. it should contain significant local structures.

**GOAL 3:** Analyse structure of company network and verify, whether it shows significant local structures.

⁶ http://graphml.graphdrawing.org/
⁷ http://jung.sourceforge.net/
Our method of network projection should be compared to other methods of projection on original networks. As a result of this goal, benefits of our work so far are revealed and proved. As additional task we should adapt our methods to other networks with temporal data. From previous goals we have mixed mode company network and projected one mode company network. As result of setting of projections we can include weights from original network and obtain various version of projected company network, also we can select only direct company to company connections from original mixed mode network for further processing. These networks are probably large and due to the nature of business environment lesser dense with local dense structures. As many positional analyses (especially partitioning techniques) are not able to process so large networks, this problem will lead us to decomposition of large networks in order to get local structures and focus on analysis of local structures. We suggest using different approaches for original subnetwork of company network (created from direct connections between companies from ORSR) and various projected one-mode company networks (built from connections via physical persons).

**GOAL 4:** Design and implement suitable method for decomposition of large company networks.

As soon as we get local structures (clusters, partitions) of network we can focus on next goal - local structure analysis. In this goal we determine possible ways of local structure analysis, its evaluation and the benefits for further use. We explore graph and matrix based techniques of positional analysis and role identifications and especially we will focus on direct and indirect approaches of block modelling on local structures and implement appropriate solutions.

**GOAL 5:** Design and implement advanced analysis of local structures in company networks.

As soon the solutions from previous goal are implemented and partitions (positions) clusters are credibly identified, we will look in compositional data and partitioning correlations using feasible knowledge discovery techniques. Result of final goal is proved correlations in data of comprehensive model data.

**GOAL 6:** Elaborate other possible uses of non-structural data in Company networks.
1.2 Book structure

In this chapter main goals have been defined. The next chapters (chapter 2 and chapter 3) contain terms and definitions from graph theory used to express formal notation of our company network and calculations used in proposed approaches. It consists from description of basic concepts of graphs, social network analysis and especially there are distinguished types of social networks and theory for social network projection.

![Logical dependencies of particular book sections](image)

Chapter 4 is dedicated to introduce used design of analyses of company networks and their specifics and the content partially contains original ideas. Next section (chapter 5) is supposed for main part of our original work where case study of Slovak company network is elaborated and in this chapter achievements of most of our goals are explained. Continually it is described here also application of designed methods and additional original ideas and methods for analysis of mixed mode company networks (Slovak company network). This content is dedicated mainly to original positional analysis, projections, dynamic analysis and their specific problems. Our work is concluded in chapter 6.