Appendix

SOFTWARE TOOLS AND APPLICATIONS FOR
DATA MINING AND CHURN MANAGEMENT

Authors’ Selection

Data mining and churn management software is generally used to identify, represent, analyze, visualize data. Especially for SNA analysis to simulate nodes (e.g. agents, organizations, or knowledge) and edges (relationships) from various types of input data (relational and non-relational), including mathematical models of social networks. The output data can be saved in external files. Various input and output file formats exist.

Various tools provide mathematical and statistical routines that can be applied to the research while ability to visually represent data is important to better understand data, network data and convey the result of the analysis. Visualization is often used as an additional or standalone data analysis method.

Our selection of tools with brief explanation and links is alphabetically summarized as follows.

IBM I2¹

*IBM i2* provides intelligence analysis, law enforcement and fraud investigation solutions. *i2* offerings deliver flexible capabilities that help combat crime, terrorism and fraudulent activity. Consisting of several tools *IBM I2* provide powerful family for intelligence, SNA and text analysis.

IBM SPSS Software²

Powerful software family for statistical analysis. *IBM SPSS* Statistics is one of the world’s leading statistical software used to solve business and research problems by means of ad-hoc analysis, hypothesis testing, and predictive analytics. Organizations use *IBM SPSS* Statistics to understand data, analyze trends, forecast and plan to validate assumptions and drive accurate conclusions.
Microsoft Excel

Part of Microsoft Office software family one of most popular spreadsheet tool with more features in every new version.

Network Overview Discovery Exploration for Excel (NodeXL)

Tool for network overview, discovery and exploration. NodeXL is a free and open Excel 2007 Add-in and C#/Net library for network analysis and visualization. It integrates into Excel 2007 and 2010 and add directed graph as a chart type to the spreadsheet and calculates a core set of network metrics and scores. Supports extracting email, Twitter, YouTube, WWW and flickr social networks.

NetworkX

Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. NetworkX (NX) is a rich integrated toolset for graph creation, manipulation, analysis, and visualization. User interface is through scripting/command-line provided by python. NX includes a large set of key algorithms, metrics and graph generators. Visualization is provided through pylab and graphviz.

Orange for Python

Open source data visualization and analysis for novice and experts. Data mining through visual programming or Python scripting. Components for machine learning. Add-ons for bioinformatics and text mining. Packed with features for data analytics.

Parc ACH

Analysis of Competing Hypotheses (ACH) is a simple model for how to think about a complex problem when the available information is incomplete or ambiguous, as typically happens in intelligence analysis. ACH is grounded in basic insights from cognitive psychology, decision analysis, and the scientific method. It helps analysts protect themselves from avoidable error, and improves their chances of making a correct judgment.
Pajek

Tool for analysis and visualization of large scale networks. A widely used program for drawing networks, Pajek also has significant analytical capabilities, and can be used to calculate most centrality measures, identify structural holes, block-model, and so on. Capable of combining data with R.

R

R environment contains several packages relevant for social network analysis: igraph is a generic network analysis package; SNA performs sociometric analysis of networks; network manipulates and displays network objects; tnet performs analysis of weighted networks, two-mode networks, and longitudinal networks; ERGM implements exponential random graph models for networks; latentnet has functions for network latent position and cluster models; degreenet provides tools for statistical modeling of network degree distributions; and etworksis provide tools for simulating bipartite networks with fixed marginals.

SQL (Structured Query Language)

Structured Query Language, special-purpose programming language designed for managing data held in a relational database management system (RDBMS), together with Python and R standard for database data manipulation

Tableau

Tableau is business intelligence software that allows anyone to connect to data in a few clicks, then visualize and create interactive, sharable dashboards with a few more. It’s easy enough that any Excel user can learn it, but powerful enough to satisfy even the most complex analytical problems. Securely sharing your findings with others only takes seconds.

Xanalys Link Explorer

Visual analytics tool combining link analysis with temporal and spatial analysis. Interactive visual analytics tool combining data acquisition and querying with link analysis, temporal analysis and spatial analysis (GIS) techniques. Integrates with other desktop applications and services such as Excel and Bing mapping
ENDNOTES

1 http://www-01.ibm.com/software/info/i2software/
2 http://www-01.ibm.com/software/analytics/spss/
4 http://nodexl.codeplex.com/
5 http://networkx.github.io/
6 http://orange.biolab.si/
7 http://www2.parc.com/istl/projects/ach/ach.html
8 http://vlado.fmf.uni-lj.si/pub/networks/pajek/
9 http://www.r-project.org/
10 http://www.tableausoftware.com/
11 http://www.xanalys.com/products/link-explorer/