Using Social Ties to Predict Missing Customer Information
Stamatis Stefanakos
DMIN 2009
Outline

- Who we are
- The data landscape in telecoms
- The call graph
- Predicting missing information from the call graph
- Conclusion
Netcetera Group AG

Netcetera Group

Zurich – Headquarters
Dr. Andrej Vckovski, CEO
Joachim Hagger, CTO
Mike Franz, Business Development
Ronnie Brunner, Corporate Development

Netcetera

Zurich, Bern, Vaduz
Dr. Simon Hefti
Managing Director

Skopje
Krume Dolnenec
Managing Director

Dubai
Jacqueline Duvoisin
Managing Director

Metaversum

Zurich
Benno Häfliger
CEO

Metaversum

Zurich
Hanspeter Gränicher
CEO

D1 Solutions

Zurich
Daniel Bareiss
CEO

Software Engineer & Application Management

Agile Methods, Java & .NET

2 Data Centers,
VISA / Mastercard certified, Swiss Banking Commission compliant

CRM

Business Engineering

Business Process Management

Business Intelligence

Consulting
Solution Design
Implementation
Operations & Support

Financial Advisory Solutions

Consulting
Tax Research
Online Tool Suite
Integrated Advisory Solutions
About D1 Solutions

Positioning
- Founded 2005 as Portfolio Company of Netcetera
- Covering both technical and business aspects of Business Intelligence
- Covering the full cycle from “data generation to decision”
- Independent from Business Intelligence products

Expertise
- Data Quality
- Data Warehousing
- Reporting / Dashboards
- Predictive Analytics / Data Mining

Customers (Selection)
- Financial Services:
  - Credit Suisse, UBS, RBS, Viseca, CSS, Zurich Financial Services, Winterthur
- Telecom:
  - Swisscom, Cablecom, Sunrise
- Utilities:
  - BKW, Axpo
- Different Industries:
  - ABB, SBB, Festo

Consulting Packages
- Business Intelligence Assessment
- Predictive Analytics & Data Mining
- Data Quality Assessment
- Process Performance Management
- Contact Center Performance
A case study in telecoms

The company:

- Number 2 in Switzerland
- Mobile & wireline
- 1.5M mobile „customers“
- Two types of mobile customers:
  - Post-paid (contract)
  - Pre-paid (refills)
The “data” landscape in a typical telecom operator

- ODS
- Billing
- Network
- External Sources

DWH

- Reporting
- Data Mining
- Promotional Campaigns
Network information (call detail records)

Every call originating from or terminating in the company’s network generates a call detail record (CDR). The CDR contains information on:

- Type of call
- Originating MSISDN, terminating MSISDN
- Date / time
- Duration
- Completion code
- Cell (Starting / Ending)
- Volume (data calls)
- Handset
- ....
Distribution of CDRs per day of the week
Distribution of CDRs per hour of day
The call graph

Is an undirected graph:
- Nodes are subscribers.
- Edges correspond to calls or short messages (SMS) between subscribers (*social ties*).

An edge exists between two nodes only if
- there are at least four connections (calls or SMS) in any direction
- there are connections in both directions

The call graph is created using four weeks of network data.
The call graph
Extracting additional information from the call graph

- We can augment the call graph with additional information about the subscribers, e.g., their age.
- The calling pattern of each subscriber reveals additional information.
- Example: The age of the social ties of a subscriber (e.g., someone who 90% of the time communicates with teenagers).
- This information can be used
  - To segment the customer base
  - To design CRM campaigns
  - To predict missing customer information
  - To identify wrong data
Age distribution of prepaid customer base
Age distribution of postpaid customer base

![Age distribution chart showing the number of postpaid customers across different age groups. The chart highlights a significant number of customers in the 20-29 and 30-39 age groups.](chart-image)
Data quality issues

Information can be missing or wrong because:
- It was never collected
- It was wrongly entered into the system
- It was wrongly aggregated
- It corresponds to a person other than the real user

These issues are usually magnified in the case of pre-paid users.
- Data is not collected because they are not needed since no contract exists
- One person buys and registers the pre-paid card but passes it on to another user
Predicting missing information or identifying wrong information through social ties

Case study: customer age

- *Show me who your friends are and I will tell you who you are*
- The age of the social ties of a customer reveals can be used to predict her real age
- We use a simple mechanism to predict the age: we average the age of all social ties
Example of the call graph induced by a single 27-year old customer
Residual error of the age prediction for postpaid customers
Mean absolute percentage error per age for postpaid customers
Concluding remarks

- Relationship information can be used to extract social ties
- Social ties can be a powerful mechanism for “creating” additional information.
- Aggregation of information through social ties can help identify missing data (or detect errors in data).
- In the case of telecoms
  - Social ties can be extracted from the network log.
  - They can help to segment the prepaid customer base (where usually information is missing).
  - They are an indicator for churn.
Contact

Dr. Stamatis Stefanakos
stamatis.stefanakos@d1solutions.ch
phone: +41 44 435 10 03 | mobile: + 41 78 634 09 24

D1 Solutions AG
Zypressenstrasse 71
CH-8040 Zürich