Mining Telecommunication Networks to Enhance Customer Lifetime Predictions

Aimée Backiel[1], Bart Baesens[1,2], and Gerda Claeskens[1]

[1] Faculty of Economics and Business, KU Leuven, Belgium
{aimee.backiel,bart.baesens,gerda.claeskens}@kuleuven.be
[2] School of Management, University of Southampton, UK

Abstract. Customer retention has become a necessity in many markets, including mobile telecommunications. As it becomes easier for customers to switch providers, the providers seek to improve prediction models in an effort to intervene with potential churners. Many studies have evaluated different models seeking any improvement to prediction accuracy. This study proposes that the attributes, not the model, need to be reconsidered. By representing call detail records as a social network of customers, network attributes can be extracted for use in various traditional prediction models. The use of network attributes exhibits a significant increase in the area under the receiver operating curve (AUC) when compared to using just individual customer attributes.

1 Introduction

Churn prediction is a common business application for classification techniques. In almost every market, companies must contend with a regular loss of customers to competition. In many markets, the level of saturation makes it much more difficult to attract entirely new customers, so the focus on customer retention is even more important. Attracting new customers is more costly than retaining existing ones [1, 2]. Longtime customers are valuable in other ways as well, including word of mouth advertising and lower cost of service. The prepaid mobile telephone segment, in particular, faces churn rates between 2 and 3% each month [1]. Accurately predicting which customers are likely to churn in advance can have a large impact on the ability to intervene and ultimately on profitability. This study, using a dataset from a major Belgian mobile provider, proposes the transformation of customer call behavior into a social network in order to enrich the data available for the classification techniques.

In this paper, two categories of churn prediction models are distinguished based on the type of attributes used for prediction: traditional models and network models. In traditional models, attributes of individual customers, such as personal information or customer behavior, are used to predict whether that individual customer will churn. The expectation is that similar customers will behave similarly with regards to churn. On the other hand, network models take into account the social network of customers, specifically calls made to other customers who have or

L. Rutkowski et al. (Eds.): ICAISC 2014, Part II, LNAI 8468, pp. 15–26, 2014.

have not churned, to predict future churners. In this case, the expectation is that people who interact with each other will behave similarly.

The experiments conducted to investigate the differences in predictive performance involve prepaid mobile customers. In this context specifically, incomplete customer information, in addition to high churn rates, necessitate advancements over the traditional prediction models. Prepaid subscriptions are to varying extents anonymous and can easily be passed to a new user without notice to the company. The outcome of this research is a prediction model built on call data available to the mobile provider which allows for timely, accurate, and interpretable predictions.

This paper is organized as follows. Following this introduction, a literature review of churn prediction and social network analysis is summarized. The data and experiment will be described in the Methodology section. Finally the results will be discussed in the Findings section. The paper concludes with a brief summary of the research.

2 Literature

2.1 Churn Prediction

Churn, in general, is defined as the loss of customers. Churn prediction is an application of classification techniques intended to predict the probability of a customer discontinuing their relationship with a company. While there are different classification techniques and a few types of attributes which can be used, the general churn prediction process is not dependent on these choices. The first step is to define churn. In some contexts, this date will be explicitly available such as a contract termination. In other contexts, such as prepaid mobile, a customer simply stops using a service. In this case, churn can be defined as a period of time without any activities on the account [3]. After the churn label is determined, the predictor variables can be used to train a classification model. This model is then applied to new data, generally current customers, to make predictions about their probability for churn. The results will form a ranking of customers from most likely to churn to least likely to churn. At this point, a percentage of customers at the top of the ranking can be contacted as part of a targeted retention campaign.

In the telecommunications domain, research has focused on mobile communications, though some landline studies have been executed as well. Mobile accounts generally include less information about the customer when compared to landline services [4], and prepaid accounts record even less customer data than postpaid accounts. In this literature review, most references use traditional prediction models, which employ customer features, such as contract type, payment amount or dates, hardware features, counts or times of calls or SMS messages. Classification techniques using local attributes include decision trees [4–8], logistic regression [2, 4–9], support vector machines [4, 8], neural networks [2, 4, 7, 9], and survival analysis [10]. Dasgupta et al.[11] use a network of customers and a relational learner to predict how churn will spread or diffuse through the network. Most churn prediction settings involve significant class skew; there is far less churn than non-churn. This should be taken into consideration as it can impact model selection, training, and evaluation. Simple classification accuracy would be less appropriate for assessing performance as a naive rule predicting all customers as non-churners would result in at least 97% accuracy given a churn rate less than 3%. Commonly used model evaluation measures in churn prediction literature include area under the ROC (receiver operating characteristic) curve, Lift which measures the percentage of top-ranked customers who are actual churners compared to the percentage of churners in the population, and assorted threshold metrics such as classification error.

2.2 Social Network Analysis

Social network analysis is based on the concept of homophily. This principle states that contact between similar people occurs at a higher rate than among dissimilar people [12]. If the contact or interaction among people in a network can be measured in some way, it can be used to make predictions instead of or in addition to the individual attributes of each person. Homophily has been used to predict links based on similar attributes and also to predict attributes based on known links [13, 8]. Social influence is a separate, but related concept explaining how individuals encourage similarity through their interactions [14]. They found that both concepts are confounded in networks and cannot be analyzed independently. For predictions, it may not be necessary to distinguish between them, but it will be difficult to interpret exactly how similarity spreads among individuals: either similar people tending to interact or interactions causing similarities.

A social network can be represented graphically where the people are nodes and some relationship between them form links or edges. Networks are not limited to people, but have actually been used to represent many different types of relationship between entities. In this literature review, applications of network analysis have been found in many areas including nursing [15], behavior adoption [16], patent classification [17], fraud detection [18], and prison system communication [19]. Pushpa and Shobha[20] and Dasgupta et al.[11] have applied network analysis to churn prediction, but their results were not compared to traditional prediction models.

Macskassy and Provost[21] have shown that simple relational learners based only on a few known class labels and the links between nodes can produce good predictions. They further suggest that traditional models and relational models should be incorporated as components into network classification systems [22]. There are two approaches for combining the different models into a single system. The approach of Macskassy and Provost begins with a local model which is used to produce prior estimates for the network. Then, a relational learner can adjust these estimates based on the links. Finally, collective inference is used when predictions for nodes are dependent on each other. This is an iterative process. The second approach works in the opposite direction. The social network can first be investigated and information collected through link mining from them can be used as attributes in a traditional model.

Different types of link features can be mined from a network. Three simple link features are mode-link, count-link, and binary link [23]. Mode-link returns the most common neighbor class. Count-link gives the number of neighbors in each class. Binary-link indicates whether there a neighbor of each class. These link features can then be used in classification models. In their study, Lu and Getoor[23] found better performance when separate models were trained for instance attributes and link features. They also determined that links between the training set and test set should be included in the network instead of creating two unconnected networks.

3 Methodology

3.1 Data and Processing

To investigate the impact of social network analysis on churn prediction, a dataset from a Belgian telco operator was analyzed.

Table 1. Data Description

Local Variables		
Account Details	Reload Information	Usage Data
Start Date	Reload count in 60 days	Numbers called in 60 days
Service Plan	Reload value in 60 days	Call time in 60 days
Trial Card	Last reload date	Numbers texted in 60 days
Language	Card swapped in 30 days	Text count in 60 days
Further Breakdowns of Local Data:		
Incoming and Outgoing, Destination Account Type		
Network Type, Day and Time of Call, Call Duration		

Network Variables	Data Type
Churn Neighbors:	Count
Churn Calling:	Time in seconds
Non-Churn Neighbors:	Count
Non-churn Calling:	Time in seconds
Out-of-Network Neighbor:	Binary
Out-of-network Calling:	Time in seconds

The data include customer information and call details from 1.4 million prepaid mobile customers from May to October 2010. The customer information includes a total of 111 local variables. Due to the nature of prepaid accounts, no personal information is available. The call detail records, over 32 million per month, include data about each call placed by a customer. These call detail records were transformed into 6 network features. Customer information and call details can be related using an anonymized phone number found in each dataset. Both sets of features are shown in Table 1.

For this study, the calls between customers are modeled as a social network. In the network, each node represents a customer and one additional node represents all non-customers. Calls between customers form single, undirected edges, weighted by the total call seconds. While text message counts were included in the customer account records, individual text messages are not present in the call detail records so they are not included in the social network. Two types of count-link features were extracted from the resulting network: counts of neighbors and the sum of call seconds with neighbors.

Churn has been defined as when a customer did not place or receive any calls for a period of 30 days. One month is short, but the cost of misclassifying a churner is higher than the a non-churner. The network nodes were labelled according to whether the customer churned during the first month. A total of six link features were then extracted from this network as shown in Table 1. Because all out-of-network neighbors are represented by a single node in the network, the Out-of-network neighbor attribute is binary, indicating whether the customer communicated with an out-of-network number or not.

The NetChurn procedure was developed to process the call detail records into social network features. The procedure has been divided into five steps for clarity and reusability. Each step is described by a short algorithm. Figure 1 visually depicts the steps and data sources in the process.

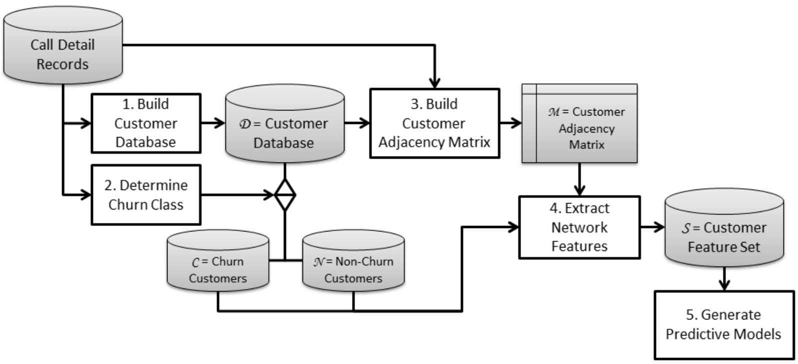


Fig. 1. Flow chart depicting the NetChurn procedure used to extract network features from Call Detail Records

Algorithm 1. Build a Customer Database from Call Detail Records

Input: Call Detail Records
Output: Customer Database
Begin with a dataset Call Detail Records CDR_m, with month m. Let $\mathcal{D}$ denote a database with customers d_1-d_n and a non-customer d_0. Each customer d_i has variables first call $first_i$ and last call $last_i$.

for *each line* $l \in CDR_1$ **do**
  if *caller* $d_i \in \mathcal{D}$ **then**
    $last_i = date_1$
  $\mathcal{D} = \mathcal{D} \cup \{d_i\}$

Algorithm 2. Determine the Churn Class for a set of Customers

Input: Call Detail Records and Customer Database
Output: Customer Database Subsets
Assume CDR_m and $\mathcal{D}$ from Algorithm 1 exist. Now let $\mathcal{D} = \mathcal{C} \cup \mathcal{N}$, such that $\mathcal{C}$ contains first month churners. The binary variable $churn_i$ indicates churn or non-churn during the entire study.

for *each line* $l \in CDR_2$-CDR_n **do**
  if *caller* $d_i \in \mathcal{D}$ **then**
    $last_i = date_l$

for *each customer* $d_i \in \mathcal{D}$ **do**
  if $last_i < churnPeriod$ **then**
    $\mathcal{C} = \mathcal{C} \cup \{d_i\}$
    $churn_i = 1$
  else if $last_i < study$ **then**
    $\mathcal{N} = \mathcal{N} \cup \{d_i\}$
    $churn_i = 1$
  else
    $\mathcal{N} = \mathcal{N} \cup \{d_i\}$
    $churn_i = 0$

Algorithm 3. Build a Customer Adjacency Matrix

Input: Call Detail Records and Customer Database
Output: Customer Adjacency Matrix
Assume CDR_m and $\mathcal{D}$ from Algorithm 1 exist. Now let $\mathcal{M}$ denote a Customer Adjacency Matrix, where each entry $m_{i,j}$ indicates the calling time between two customers d_i and d_j.

for *each line* $l \in CDR_1$ **do**
  if *caller* d_i *and receiver* $d_j \in \mathcal{D}$ **then**
    $m_{i,j} = m_{i,j} + time_l$
  else if *caller* $d_i \in \mathcal{D}$ **then**
    $m_{0,i} = m_{0,i} + time_l$

Algorithm 4. Extract Network Features

Input: Customer Adjacency Matrix and Customer Database Subsets
Output: Customer Network Features
Assume $\mathcal{D} = \mathcal{C} \cup \mathcal{N}$ from Algorithm 2 and $\mathcal{M}$ from Algorithm 3 exist.
Now let $\mathcal{S}$ denote a set of customers s_1-s_n, each with six network features: Churn neighbors $countC_i$, Churn calling $timeC_i$, Non-churn neighbors $countNC_i$, Non-churn calling $timeNC_i$ Out-of-network neighbor $hasOut_i$, and Out-of-network calling $timeOut_i$

```
for each customer d_j D do
    for each customer d_i D | i < j do
        if d_i = 0 then
            hasOut_j = 1
            timeOut_j = m_{0,j}
        else if d_i ∈ C then
            countC_j = countC_j + 1
            timeC_j = timeC_j + m_{i,j}
        else if d_i ∈ N then
            countNC_j = countNC_j + 1
            timeNC_j = timeNC_j + m_{i,j}
```

As discussed in Section 2.2, social network analysis is based on the concept of homophily, that is, that similar people tend to interact more than dissimilar people. Easley and Kleinberg[24] explain one possible test for homophily is to compare the actual fraction of cross-gender edges to the expected cross-gender edges in a random network. If the links in the social network built in this study had been assigned randomly, the expected proportion of edges between churners and non-churners would be 0.3384. In the actual network, the proportion of cross-gender edges is only 0.1391. A t-test showed a statistically significant difference with p-value <0.001. Homophily implies that the network attributes will enhance churn prediction. Figure 2 depicts a subset of the network where the density of links between churners can be seen. This image was generated using Pajek - Program for Large Network Analysis (http://pajek.imfm.si/doku.php, accessed 11 September 2013).

3.2 Experimental Setup

After the data processing was complete, as experiment was designed to compare the use of local attributes versus network attributes. Two types of models were tested: logistic regression and Cox proportional hazards regression (Cox PH). Three variable sets were tested: local attributes, network attributes, and a combination of both. All models were estimated in SAS, which is accomplished with an iterative maximization process [25]. Logistic regression was chosen as an established classification technique. In many applications, logistic regression offers good predictive performance, understandable models, and interpretable

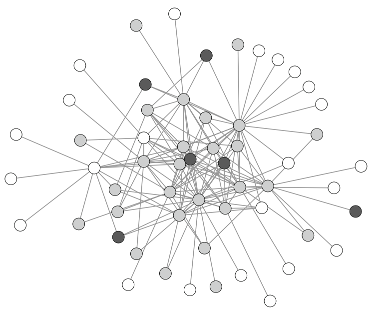


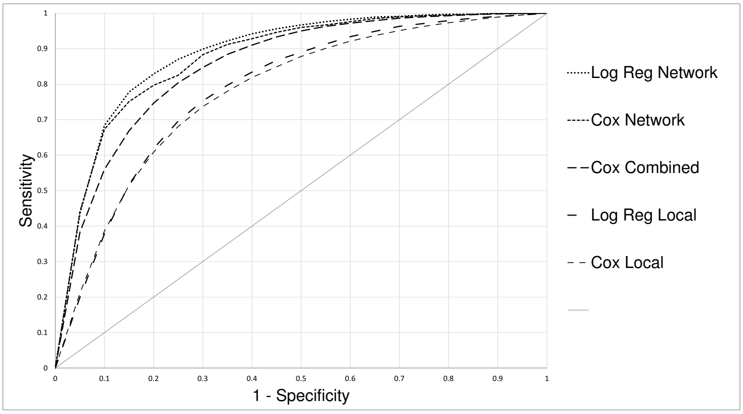
Fig. 2. Cluster of churners in the network. Key: non-churners (white), first month churners (light grey), later churners (dark grey).

log odds ratio to assess the explanatory variables. On the other hand, the Cox proportional hazards model was selected because it allows for time-to-event predictions, in this case, a likely time of churn can be estimated for each customer, rather than a binary churn or non-churn label. To compare this with logistic regression, it is necessary to train a logistic regression model for each time period, while one Cox PH model can make predictions for all time periods. Similar to the log odds ratio, the Cox PH model results in hazard ratios indicating how the probability of churn changes when each explanatory variables changes value.

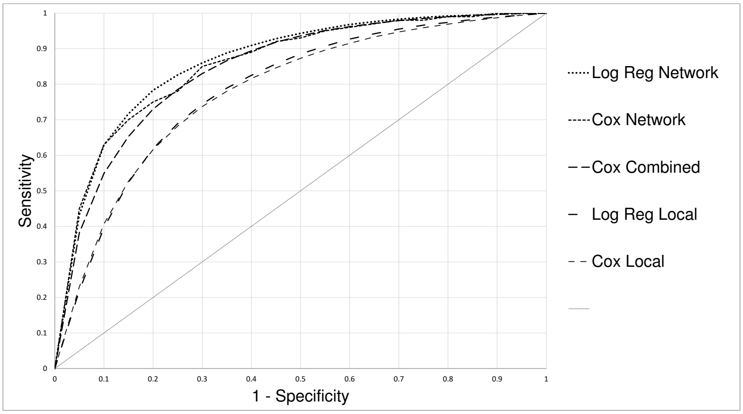
The models were trained on a subset of 70% of the data. A holdout test set or the remaining 30% was used for validation. Feature selection, based on the Wald test, reduced the local attribute feature set. The Bonferroni correction was used to avoid false positives by lowering the p-value to <0.0001. Ultimately, 39 local attributes were included in the local model. In the network feature set, the out-of-network neighbor binary was not significant, so the remaining five network attributes were included. In the combined model, 41 features were found to be significant. The time interval for predictions was set to one month, since logistic regression requires a different model for each time interval.

4 Findings

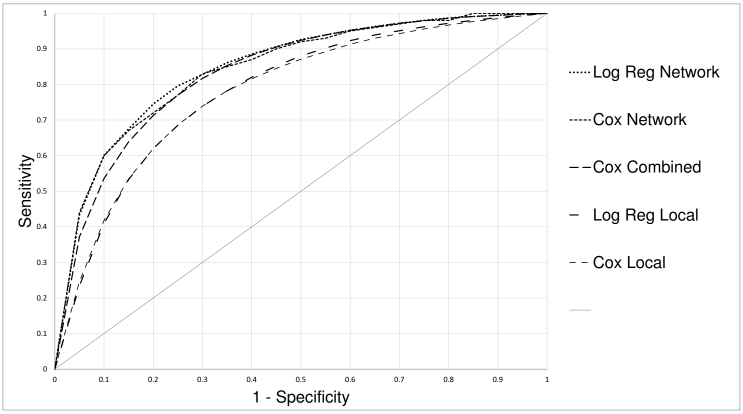
As a first visual evaluation, ROC curves have been plotted for three months of predictions in Figure 3. ROC curves are graphical representations of the true positive rate versus the false positive rate. A perfect classification is the point (0,1)—0% false positives and 100% true positives—on the graph, so the closer an ROC curve is to that point, the better the model is [26]. The value of interest in regression models is a probability. In order to label a customer as a predicted churner or non-churner, a cut-off point is used to separate the classes. Probabilities above this threshold will be assigned as churners and those below it will be assigned as non-churners. Each point on the ROC curve corresponds to one such cut-off score in the full range between 0.0 and 1.0.



(a) Month 1 ROC Curves



(b) Month 2 ROC Curves



(c) Month 3 ROC Curves

Fig. 3. Comparison of Prediction Models

Models can be quantitatively compared by calculating the Area Under the ROC Curves (AUC). The differences between models was tested using the test of DeLong et al.[26]. In the first two months, the Cox PH Model with network attributes is not significantly better than the Cox PH Model with combined attributes. All other models are significantly different. In the third month, the differences in AUC values for all models were found to be significant. The AUC values for the five models over three months are displayed in Table 2. In all three months, it can be seen that models with network features included do result in greater AUC values than those with only local attributes. However, the differences between network and local models are most prominent in the first month of predictions. While the network models continue to outperform the local models, the AUC values for the local models remain relatively constant while the network models decline over time.

When comparing the models using only network features to those including customer attributes, far fewer features are included in network models. These less complex models are more easily interpreted in addition to their superior performance. The contribution of each feature to the prediction can be analyzed and this information leads to a better understanding of the influences on churn.

Table 2. AUC Values for Month 1, Month 2, and Month 3

Month 1	**Features**	**AUC**	**Error**
Log Reg. Network	6	**0.8836**	0.0014
Cox PH Network	5	0.8735[1]	0.0015
Cox PH Combined	41	0.8691[1]	0.0014
Log Regression Local	29	0.7872	0.0020
Cox PH Local	39	0.7792	0.0020
Month 2	**Features**	**AUC**	**Error**
Log Reg. Network	6	**0.8601**	0.0012
Cox PH Network	5	0.8543[2]	0.0012
Cox PH Combined	41	0.8553[2]	0.0011
Log Regression Local	29	0.7850	0.0015
Cox PH Local	39	0.7794	0.0015
Month 3	**Features**	**AUC**	**Error**
Log Reg. Network	6	0.8436	0.0011
Cox PH Network	5	0.8425	0.0011
Cox PH Combined	41	**0.8462**	0.0010
Log Reg. Local	29	0.7843	0.0012
Cox PH Local	39	0.7794	0.0012

Comparison of churn prediction models based on different feature sets. For each month, the greatest AUC is highlighted. [1] [2] No significant difference.

5 Summary

This study was designed to investigate the differences in predicting customer churn by using features based on customers individually or based on the network formed by calls among customers. When evaluating the models using AUC values, it has been shown that network features do offer improved predictions when compared to local features, though the differences are more pronounced on sooner predictions than later. This suggests that network information may be more dynamic, reflecting more current changes in customer behavior. According to these findings, valuable information is present in the social networks of customers. Taking advantage of this information can improve churn prediction models and the retention campaigns designed to reduce churn.

A more practical contribution of this work is a business-oriented approach, implemented in Java, to process one month of company data and make predictions for the next few weeks or months. Using this approach, companies can employ an ongoing monthly targeted retention campaign which identifies customers at risk for churn dynamically.

Acknowledgements. —This research was made possible with support of the Odysseus program (Grant B.0915.09) of the Fund for Scientific Research–Flanders (FWO).

References

1. Bersen, A., Smith, S., Thearling, K.: Building Data Mining Applications for CRM. McGraw-Hill, New York (2000)
2. Verbeke, W., Martens, D., Mues, C., Baesens, B.: Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Systems with Applications 38, 2354–2364 (2011)
3. Owczarczuk, M.: Churn models for prepaid customers in the cellular telecommunication industry using large data marts. Expert Systems with Applications 37, 4710–4712 (2010)
4. Huang, B., Kechadi, M.T., Buckley, B.: Customer churn prediction in telecommunications. Expert Systems with Applications 39, 1414–1425 (2012)
5. Lima, E., Mues, C., Baesens, B.: Domain knowledge integration in data mining using decision tables: Case studies in churn prediction. Journal of Operational Research Society 60(8), 1096–1106 (2009)
6. Risselada, H., Verhoef, P., Bijmolt, T.: Staying power of churn prediction models. Journal of Interactive Marketing 24, 198–208 (2010)
7. Verbeke, W., Dejaeger, K., Martens, D., Hur, J., Baesens, B.: New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. European Journal of Operational Research 218, 211–229 (2012)
8. Zhang, X., Zhu, J., Xu, S., Wan, Y.: Predicting customer churn through interpersonal influence. Knowledge-Based Systems 28, 97–104 (2012)
9. Verbraken, T., Verbeke, W., Baesens, B.: A novel profit maximizing metric for measuring classification performance of customer churn prediction models. IEEE Transactions on Knowledge and Data Engineering 25(5), 961–973 (2013)

10. Wong, K.K.K.: Using cox regression to model customer time to churn in the wireless telecommunications industry. Journal of Targeting, Measurement, and Analysis for Marketing 19(1), 37–43 (2011)
11. Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A., Joshi, A.: Social ties and their relevance to churn in mobile telecom networks. In: EDBT 2008 Proceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology, Nantes, France, pp. 668–677. ACM (March 2008)
12. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Annual Review of Sociology 27, 415–444 (2001)
13. Rhodes, C.J., Jones, P.: Inferring missing links in partially observed social networks. Journal of Operational Research Society 60(10), 1373–1383 (2009)
14. Shalizi, C.R., Thomas, A.C.: Homophily and contagion are generically confounded in observational social network studies. Sociological Methods and Research 40(2), 211–239 (2011)
15. Pow, J., Gayen, K., Elliott, L., Raeside, R.: Understanding complex interactions using social network analysis. Journal of Clinical Nursing 21, 2772–2779 (2012)
16. Mertens, F., Saint-Charles, J., Mergler, D.: Social communication network analysis of the role of participatory research in the adoption of new fish consumption behaviors. Social Science and Medicine 75(4), 643–650 (2012)
17. Liu, D.-R., Shih, M.-J.: Hybrid-patent classification based on patent-network analysis. Journal of the American Society for Information Science and Technology 62(2), 246–256 (2010)
18. Chiu, C., Ku, Y., Lie, T., Chen, Y.: Internet auction fraud detection using social network analysis and classification tree approaches. International Journal of Electronic Commerce 15(3), 123–147 (2011)
19. Hancock, P.G., Raeside, R.: Analysing communication in a complex service process: an application of social network analysis in the scottish prison service. Journal of Operational Research Society 61, 265–274 (2009)
20. Pushpa, Shobha, G.: An efficient method of building the telecom social network for churn prediction. International Journal of Data Mining and Knowledge Management Proces 2(3), 31–39 (2012)
21. Macskassy, S.A., Provost, F.: A simple relational classifier. In: Proceedings of the Second Workshop on Multi-Relational Data Mining (MRDM-2003) at KDD-2003, pp. 64–76 (2003)
22. Macskassy, S.A., Provost, F.: Classification in networked data: A toolkit and a univariate case study. Journal of Machine Learning Research 8(2), 935–983 (2007)
23. Lu, Q., Getoor, L.: Link-based classification using labeled and unlabeled data. In: Proceedings of the ICML Workshop on The Continuum from Labeled to Unlabeled Data, Washington, DC, USA, ICML (2003)
24. Easley, D., Kleinberg, J.: Networks, Crowds, and Markets. Cambridge University Press, Cambridge (2010)
25. Allison, P.D.: Survival Analysis Using SAS: A Practical Guide, 2nd edn. SAS Institute Inc., Cary (2010)
26. DeLong, E., DeLong, D., Clarke-Pearson, D.: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. International Biometric Society 44(3), 837–845 (1988)