A Content-Based Social Network Study of Evliyâ Çelebi's *Seyahatnâme-Bitlis Section*

Ceyhun Karbeyaz, Ethem F. Can, Fazli Can and Mehmet Kalpakli

Abstract Evliyâ Çelebi, an Ottoman writer, scholar and world traveler, visited most of the territories and also some of the neighboring countries of the Ottoman Empire in the seventeenth century. He took notes about his trips and wrote a 10-volume book called *Seyahatnâme (Book of Travels)*. In this paper, we present two methods for constructing social networks by using textual data and apply it to *Seyahatnâme-Bitlis Section* from book IV. The first social network construction method is based on proximity of co-occurence of names. The second method is based on 2-pair associations obtained by association rule mining by using sliding text blocks as transactions. The social networks obtained by these two methods are validated using a Monte Carlo approach by comparing them with the social network created by a scholar-historian.

1 Introduction

Evliyâ Çelebi; a seventeenth century Ottoman writer, scholar, and world traveler (born on 1611, died circa 1682); visited most of the territories and also some of the neighboring countries in Africa, Asia and Europe of the Ottoman Empire over a

C. Karbeyaz · E. F. Can · F. Can (✉)
Department of Computer Engineering, Bilkent University, 06800 Ankara, Turkey
e-mail: canf@cs.bilkent.edu.tr

C. Karbeyaz
e-mail: karbeyaz@cs.bilkent.edu.tr

E. F. Can
e-mail: efcan@cs.bilkent.edu.tr

M. Kalpakli
Department of History, Bilkent University, 06800 Ankara, Turkey
e-mail: kalpakli@bilkent.edu.tr

period of 40 years. His work *Seyahatnâme (The Book of Travels)* is known by its distinguished style and detailed descriptions of people and places that he visited during his long journeys [1]. UNESCO has declared 2011 as the anniversary year of Evliyâ Çelebi on the four hundredth anniversary of his birth. This provides an additonal motivation for this study.

One of the methods we present in this study is based on association rules, which are the derived relations between the items of a dataset. Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of items with size n and $T = \{t_1, t_2, \ldots, t_m\}$ is set of transactions (in market data analysis a transaction involves the group of items purchased together) with size m. Then an association rule is shown as $X \Rightarrow Y$ where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. In a related work Raeder and Chawla [2] model a store's product space as a social network using association rules. Their work shows the use of social networks for market basket analysis.

The important aspects of this study can be summarized as follows. We present two different methods for constructing social networks from textual data and apply them to *Seyahatnâme-Bitlis Section* from book IV. For this purpose, we use the text in transcribed form [3]. We employ the social network created by a human expert as the ground truth, and assess the effectiveness of the methods by comparing the generated network structure with that of the ground truth. Finally, we use a Monte Carlo approach and show that the social network structures obtained by our methods are significantly different from random, i.e., are not by chance and hence valid.

2 Methods and Measuring Effectiveness

In this paper, we present two social network construction methods. These are the text proximity-based method (ProxiBM) and the association rule-based method (RuleBM). Both methods are based on co-occurrence of names in close proximity within a text block. For determining text blocks with a meaningful cohesive context we use two approaches. In the first blocking approach we use the paragraph information provided in the transcribed text. We manually identified and tagged 164 paragraphs. Each paragraph is used as a block. In the second method we employ a sliding window-based blocking approach (see Fig. 1).

In ProxiBM, edges for the undirected graph of social network are derived by creating a link between every character that appear in the same paragraph within a close word proximity (using a threshold). The proximity threshold between any two names is varied between 5 words and 500 words in steps. This approach is inspired by the use of term closeness as an indicator of document relevance [4].

In RuleBM we use the sliding text window for blocking and treat each block as a transaction where the existing names correspond to shopping items. We use the 2-pair association rules as relational edges of the social network by using the Apriori algorithm [5]. We employ different support threshold values and repeat

Fig. 1 Sliding window-
based blocking:l total text
length, b block size
$(0 < b \le l)$, s step size,
nb number blocks, $nb = 1 +$
$\lceil \frac{(l-b)}{s} \rceil$ for $0 < s \le b$, $nb =$
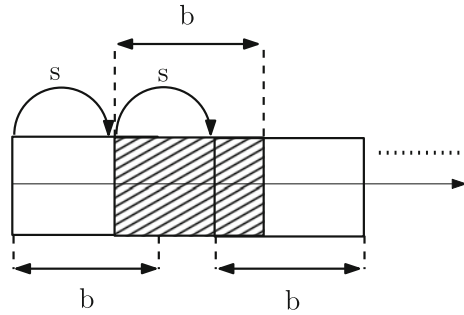$1 + \lfloor \frac{(l-b)}{s} \rfloor$ for $s > b$

Table 1 Performance results of ProxiBM over paragraphs for different proximity threshold (Θ) values in terms of no. of words

Measure	$\Theta = 5$	$\Theta = 10$	$\Theta = 25$	$\Theta = 50$	$\Theta = 100$	$\Theta = 250$	$\Theta = 500$
Precision	0.47	0.52	**0.54**	0.49	0.48	0.46	0.45
Recall	0.16	0.39	0.66	0.70	0.70	**0.71**	**0.71**
F-measure	0.24	0.44	**0.59**	0.58	0.57	0.56	0.56

the blocking operation for different block and step sizes in order to find the best performing parameters. The agreement between automatically constructed social networks and the manually constructed (actual) social network is measured by precision, recall, and the F-measure [6, pp. 142–144].

3 Experimental Results

In the experiments, both methods are tested in various conditions in order to find their best matching configuration to the ground truth. For ProxiBM we use various proximity threshold values and for RuleBM different blocksize, step size and support threhold values. Table 1 shows the precision, recall and F-measure results of ProxiBM for different proximity threshold values. The best configuration (the highest F-measure value) for this method is observed when proximity threshold is 25 words.

A similar experiment is done for RuleBM. Association rules correspond to frequencies of these 2-pair items (character names) appearing together in different transactions (blocks). They are derived from the blocks for different support thresholds ranging from 5 to 20%. The precision, recall and F-measure results of the RuleBM for the best configuration which is 500 words, stepsize: 300 words and support threshold: 5% can be seen in Table 2 among with other support threshold values.

Table 2 Performance results of RuleBM for blocksize: 500 words, stepsize: 300 words and for different support threshold (β) values

Blocksize	Stepsize	Measure	$\beta = 5\%$	$\beta = 10\%$	$\beta = 15\%$	$\beta = 20\%$
500	300	Precision	0.24	**0.31**	0.20	*
		Recall	**0.38**	0.07	0.01	*
		F-measure	**0.29**	0.11	0.01	*

* No association rules are found for that configuration.

Automatically generated social networks are further tested to understand if they are significantly different from random. For this purpose Monte Carlo experiments are performed [7]. In all Monte Carlo experiments, we generate a random version of the social network which is being evaluated 1000 times and measure the average F-measure values. In order to achieve this the Erdos-Renyi random network generation algorithm is used [8]. Monte Carlo results show that both methods with proper parameters generate networks which are significantly different from random.

4 Conclusion and Future Work

We present two methods ProxiBM and RuleBM for constructing social networks by using textual data and apply it to *Seyahatnâme-Bitlis Section* from book IV. The experimental results show that the networks created by ProxiBM show a higher similarity to the manually created social network than those of RuleBM. However, the disadvantage of ProxiBM is that it requires more focused (cohesive) blocks obtained from paragraphs. On the other hand, RuleBM is more flexible since it simply exploits blocks obtained from a sliding text window.

It is possible to obtain a better performance with RuleBM if we use contextually meaningful sliding text blocks: For the construction of such cohesive units we may use an automatic text segmentation method [9].

References

1. Dankoff, R.: Evliyâ Çelebi in Bitlis: The Relevant Section of the Seyahatn âme. E.J. Brill, Netherlands (1990)
2. Raeder, T., Chawla, N.V.: Modeling a store's product space as a social network. In: Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining, pp. 164–169, IEEE Computer Society, Washington DC, USA (2009)

3. Kahraman, S.A., Dağlı, Y.: Evliyâ Çelebi Seyahatnâmesi IV. Kitap. Yapı Kredi Yayınları, İstanbul (2003)
4. Hawking, D., Thistlewaite, P.: Relevance weighting using distance between term occurrences. Technical Report , The Australian National University, Canberra (1996)
5. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases. VLDB '94, pp. 487–499, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1994)
6. Manning, C.D., Raghavan, P., Schütze H.: Introduction to Information Retrieval. Cambridge University Press, New York (2008)
7. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs, NJ (1988)
8. Erdős P., Rényi, A.: On random graphs. I. Publ. Math. Debrecen 6:290–297 (1959)
9. Hearst, M.A.: Texttiling: segmenting text into multi-paragraph subtopic passages. Comput. Linguist. **23**(1), 33–64 (1997)