

# Understanding social networks using Formal Concept Analysis

Václav Snášel      Zdeněk Horák      Ajith Abraham  
VSB Technical University Ostrava, Czech Republic  
{vaclav.snasel, zdenek.horak.st4}@vsb.cz  
Center of Excellence for Quantifiable Quality of Service  
Norwegian University of Science and Technology, Norway  
{ajith.abraham@ieee.org}

## Abstract

*Social networks are very popular nowadays and the understanding of their inner structure seems to be promising area. Several approaches for the social network structure visualization has been proposed and the typical problem is the schematic value of such visualization and the computational complexity of their analysis. This paper propose a method by using the relation between objects, which is reduced and uses as an input to Formal Concept Analysis methods. The proposed method attempts to deal with both mentioned problems.*

## 1 Introduction

The social network is a structure made of **nodes** (often representing persons), which are **tied together** by some kind of relationship (representing for example friendship).

The availability of this type of data has grown significantly due to the immense growth of internet access, because many online systems based on user interaction could be viewed as social networks. Visualization of these networks is not only attractive, but can be also the basis for discovering terrorist networks or making some managerial decisions.

Thus are several types of visualization schemes and for a good overview covering both computer based and hand drawn methods, please consult [3]. The main purpose of network visualization is to give a quick overview of the network structure and also to show what is not apparent directly from the data. There are also some applications for realtime visualization and navigation created Yun and Boqin [6]). [12] noticed the limitation of classical graphs when using them for social networks visualizations.

As the social network is based on the relationships, we may think of using Formal Concept Analysis, a well established method for analysis of tabular data [5], [4], [13]. Due

to the massive dimensions of the real-world social networks, we may run into computational complexity troubles and the illustrative value is often a puzzle. Several research works are dealing with this problem [2]. Rice et al. [10] used a clustering approach for generating parts of the concept lattice.

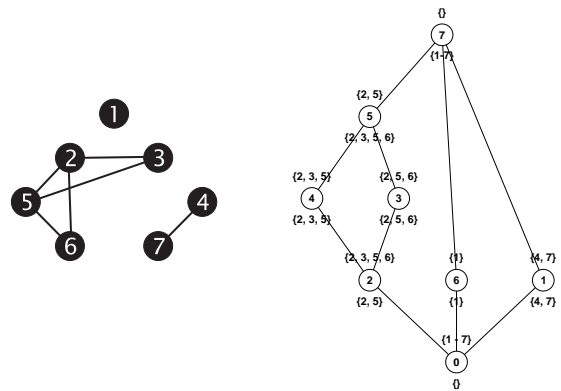


Figure 1. Social network graph and the corresponding lattice before reduction

The proposed approach differs in using a **combination of two methods**. We reduce the input data using matrix factorization methods before analyzing them. In this paper, a Singular Value Decomposition method is used, but also other methods may be considered (see [7] for Non-negative matrix decomposition). Another aspect worth mentioning is that we can also consider the strength of relationship between subjects. This leads to a matrix consisting of real numbers. For performance and clarity reasons, we considered the binary case only (that means the subjects are or are not related), but the proposed approach is relatively general and can be used in similar manner for fuzzy relations [1].

The paper is organized as follows. In Section 2, we in-

roduce the basics of formal concept analysis and singular value decomposition followed by an illustrative example in Section 3. Experiment results are reported in Section 4 and some conclusions are provided towards the end.

## 2 Preliminaries

### 2.1 Formal concept analysis

Formal Concept Analysis (shortly FCA, introduced by **Rudolf Wille** in 1980) is based on the understanding of the world in terms of objects and attributes. It is assumed that a relation exists to connect objects to the attributes they possess.

A **formal context**  $C = (G, M, I)$  is a triplet consisting of two sets,  $G$  and  $M$ , with  $I$  in relation to  $G$  and  $M$ . The elements of  $G$  are defined as objects and the elements of  $M$  are defined as attributes of the context. In order to express that an object  $g \in G$  is related to  $I$  with the attribute  $m \in M$ , we record it as  $gIm$  or  $(g, m) \in I$  and read that object  $g$  has the attribute  $m$ .

For a set  $A \subseteq G$  of objects we define  $A' = \{m \in M \mid gIm \text{ for all } g \in A\}$  (the set of attributes common to the objects in  $A$ ). Correspondingly, for a set  $B \subseteq M$  of attributes we define  $B' = \{g \in G \mid gIm \text{ for all } m \in B\}$  (the set of objects which have all attributes in  $B$ ).

A **formal concept** of the context  $(G, M, I)$  is a pair  $(A, B)$  with  $A \subseteq G$ ,  $B \subseteq M$ ,  $A' = B$  and  $B' = A$ . We call  $A$  the extent and  $B$  the intent of the concept  $(A, B)$ .  $\mathcal{B}(G, M, I)$  denotes the set of all concepts of context  $(G, M, I)$  and forms a complete lattice (so called **Galois lattice**). Reader may consult [5] for more technical details and [9] for the connection between graphs and FCA.

### 2.2 Singular Value Decomposition

Singular value decomposition (SVD) is well-known because of its application in information retrieval. SVD is especially suitable in its variant for sparse matrices [8].

**Theorem 1.** *Let  $A$  is an  $m \times n$  rank- $r$  matrix. Let  $\sigma_1 \geq \dots \geq \sigma_r$  be the eigenvalues of a matrix  $\sqrt{AA^T}$ . Then there are orthogonal matrices  $U = (u_1, \dots, u_r)$  and  $V = (v_1, \dots, v_r)$ , whose column vectors are orthonormal, and a diagonal matrix  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$ . The decomposition  $A = U\Sigma V^T$  is called singular value decomposition of matrix  $A$  and numbers  $\sigma_1, \dots, \sigma_r$  are singular values of the matrix  $A$ . Columns of  $U$  (or  $V$ ) are called left (or right) singular vectors of matrix  $A$ .*

Now we have a decomposition of the original matrix  $A$ . We have at most  $r$  non-zero singular numbers, where rank  $r$  is the smaller of the two matrix dimensions. Because

the singular values usually fall quickly, we can take only  $k$  greatest singular values and corresponding singular vector coordinates and create a  $k$ -reduced singular decomposition of  $A$ .

Let us have  $k$ ,  $0 < k < r$  and singular value decomposition of  $A$ . We call  $A_k = U_k E_k V_k^T$  a  $k$ -reduced singular value decomposition (rank- $k$  SVD).

**Theorem 2.** (Eckart-Young) *Among all  $m \times n$  matrices  $C$  of rank at most  $k$ ,  $A_k$  is the one, that minimises  $\|A_k - A\|_F^2 = \sum_{i,j} (A_{i,j} - C_{w,j})^2$ .*

Briefly said, SVD allows us to decompose one matrix into several others. By multiplying them back we can obtain the original matrix. Another choice is to remove some part of decomposed matrices before the multiplication, which will give us matrix similar to the original one.

Note: From now on, we will assume rank- $k$  singular value decomposition when speaking about SVD.

## 3 Illustrative Example

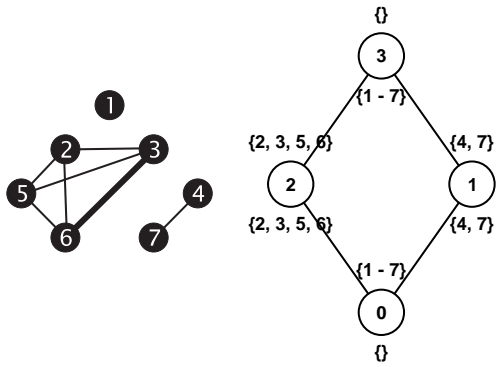
User	U1	U2	U3	U4	U5	U6	U7
U1	(1)	0	0	0	0	0	0
U2	0	1	1	0	1	1	0
U3	0	1	1	0	1	(0)	0
U4	0	0	0	1	0	0	1
U5	0	1	1	0	1	1	0
U6	0	1	(0)	0	1	1	0
U7	0	0	0	1	0	0	1

**Table 1. Social network incidence relation**

Social network can be modeled of as a set of subjects, in which some of them has some relationship with others. This can be formalized as a classical mathematical relationship and visualized for example as a undirected graph.

Table 1 depicts a a very simple system with 7 users. Users U4 and U7 know each other. User U1 does not know any other user. Users U2, U3, U5 and U6 knows each other with one exception and user U3 does not know user U6. The same table can be used as an incidence relation for graph, which is illustrated in the left part of Figure 1.

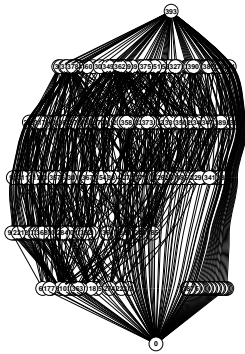
Now we can employ methods of FCA to illustrate the inner structure of the data. Concept lattice formulation is illustrated on the right side of Figure 1. Each node represents one concept from the lattice. The extent of the concept is depicted on the top of every node and the corresponding intents are noted down below the nodes. Both the graph and the lattice are quite illustrative, but if we consider bigger systems consisting of hundreds of users, the illustration becomes very chaotic (Please see Figure 3).



**Figure 2. Social network graph and corresponding lattice after reduction**

At this point, we use the binary matrix factorization and as illustrated in [11], we reduce the binary table using SVD. At some setting of rank, the same table is obtained with three cells changed. These cells are highlighted by parenthesis in the aforementioned original table.

As evident, the reduction has added relationship between users 3 and 6. That is quite natural because if two users are connected to the same users, they will probably – in some way – be connected to themselves. The reduction also changed the relationship of user 1 to itself. Modified graph is illustrated in Figure 2. In this case, the reduction created a graph clique.



**Figure 3. Concept lattice from original data**

On the right side of Figure 2, the modified concept lattice is illustrated, which can be computed faster and well reflects the situation, that in the original data have contained two interesting groups.

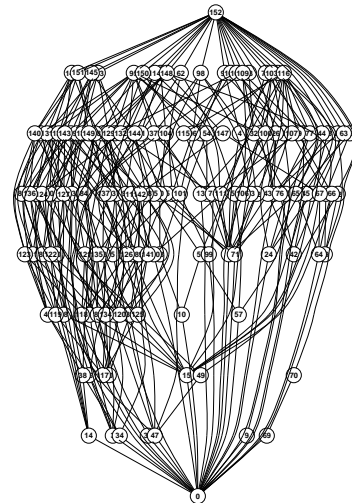
Another interesting aspect is from the diagonal of the

matrix ( represents the incidence relation), which is all unity. This can be interpreted as everybody is in a relationship with themselves. This is quite natural and also has practical advantage, for example: group of users which know themselves can share the same attribute signature.

## 4 Experiments

We may consider data from some real system containing relations between hundreds of users. In such case, we do not need the exact results, because they will be uncomputable using standard FCA methods and it will be unmanageable to have a detailed analysis at the results. The usage of our proposed method is sound from this point of view.

As a social network, we considered the data obtained by the Federal Energy Regulatory Commission’s investigation of Enron corporation (so called Enron corpus<sup>1</sup>). We have taken the e-mail recipients as persons in the constructed social network and created a relationship between them if they have exchanged some predefined amount of messages. For better illustration and computational processing, we have selected only users, which were in contact with several other users. From the same reason, we had to select only some amount of them. More precisely, we have obtained a binary square matrix with 150 rows/columns, with ones on the diagonal and more then 7 ones in a row. Visualization of this matrix using FCA as concept lattice is depicted in Figure 3.



**Figure 4. Reduced concept lattice (rank 15)**

Using the procedure illustrated in Section 3, we can reduce the input matrix and compute the reduced concept lat-

<sup>1</sup>see [http://arg.vsb.cz/arg/Enron\\_Corpus/default.aspx](http://arg.vsb.cz/arg/Enron_Corpus/default.aspx) for more details

tices. Using different ratios of reduction, we can get several lattices and some of them are illustrated in figures 4, 5 and 6. The interpretation of the created concepts depends on the content of the communication itself and we are not dealing with this in this paper.

## 5 Conclusions

We have proposed a novel approach to overcome some practical issues when dealing with analysis and visualization of large scale social networks data. Our preliminary results indicate that the approach is helpful for solving these issues but only further detailed experiments can prove this.

As future work, we would like to focus on bigger networks and interpretation of the analysis within some limited area.

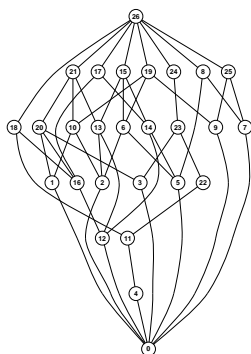


Figure 5. Reduced concept lattice (rank 5)

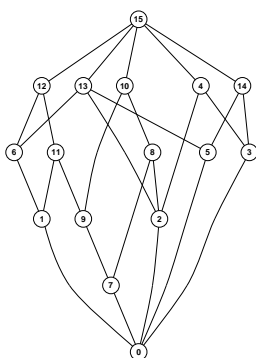


Figure 6. Reduced concept lattice (rank 4)

## References

- [1] R. Bělohávek, V. Vychodil: What is a fuzzy concept lattice, Proceedings of the CLA, pp 7–9 (2005)
- [2] R.J. Cole, P.W. Eklund: Scalability in Formal Concept Analysis, Computational Intelligence, vol 15, 11–27 (1999)
- [3] L. C. Freeman: Visualizing social networks, Journal of social structure 1 (2000)
- [4] L. C. Freeman, D. R. White: Using Galois Lattices to Represent Network Data, Sociological Methodology, vol 23, pp 127 – 146 (1993)
- [5] B. Ganter, R. Wille: Formal Concept Analysis: Mathematical Foundations, Springer-Verlag, New York (1997)
- [6] J. Heer, D. Boyd, Vizster: Visualizing Online Social Networks, Proceedings of the 2005 IEEE Symposium on Information Visualization, pp. 33–40 (2005)
- [7] D. Lee, H. Seung.: Learning the parts of objects by non-negative matrix factorization, Nature, 401, 788–791 (1999)
- [8] T. Letsche, M. Berry, S. Dumais.: Computation methods for intelligent information access, Proceedings of the 1995 ACM/IEEE Supercomputing Conference (1995)
- [9] M. Liquiere: Some links between Formal Concept Analysis and graph mining, Mining Graph Data, Wiley-Interscience (2006)
- [10] M. D. Rice, M. Siff: Clusters, Concepts and Pseudometrics, Electronic Notes in Theoretical Computer Science (2002)
- [11] V. Snášel, M. Polovinčák, H. M. Dahwa, Z. Horák: On concept lattices and implication bases from reduced contexts, Supplementary Proceedings of the 16th International Conference on Conceptual Structures, ICCS 2008, pp 83 – 90 (2008)
- [12] F. B. Vidas, J. Donath: Social Network Visualization: Can We Go Beyond the Graph?
- [13] Z. Yun, F. Boqin: Effective Browsing of Personal Tag Space in Social Tagging Systems, IEEE International Conference on Information Reuse and Integration (2008)