
The analysis of split graphs in social networks based on the k -cardinality assignment problem

Ivan Belik

Norwegian School of Economics,
Helleveien 30, 5045 Bergen, Norway
Email: Ivan.Belik@nhh.no

Abstract: In terms of social networks, split graphs correspond to the variety of interpersonal and intergroup relations. In this paper, we analyse the interaction between cliques (socially strong and trusty groups) and independent sets (fragmented and non-connected groups) as the basic components of any split graph. Based on the semi-Lagrangian relaxation for the k -cardinality assignment problem, we show the way of minimising socially risky interactions between cliques and independent sets within social networks.

Keywords: social networks; split graphs; cliques; independent sets; assignment problem; semi-Lagrangian relaxation; network science.

Reference to this paper should be made as follows: Belik, I. (2016) 'The analysis of split graphs in social networks based on the k -cardinality assignment problem', *Int. J. Network Science*, Vol. 1, No. 1, pp.53–62.

Biographical notes: Ivan Belik is a Research Scholar at the Norwegian School of Economics (NHH). He received his degree in Computer Science from the Southern Illinois University Carbondale (SIUC). He is a Fulbright Alumni (2009-2011, USA), Golden Key International Honour Society member (since 2010, USA), and the member of the International Association of Engineers (IAENG).

1 Introduction

The understanding of how social networks are currently forming, what kinds of relations exist and possible ways to formalise, predict and manipulate these networks through their internal mechanisms and structures are important in many socio-economic settings. Socially generated networks are different from any other types of networks in terms of their internal structure, interaction mechanisms and tools employed to analyse them. Jackson and Rogers (2007) consider large-scale social networks as dynamic models in trying to understand how random social networks are. They argue that different situations where social network structures "...play a central role include scientific collaborations among academics, joint research ventures among firms, political alliances, trade networks, the organization of intrafirm management, the sharing of information about job

opportunities...”. Analysing the formation of social networks they present a dynamic model where nodes in social networks form relations in two ways:

- 1 randomly
- 2 by searching locally through the structure.

The main feature of any social network is its natural intention for the formation of internal communities and modules. The analysis of social networks’ modularity is one of the most challenging problems in the area of social network analysis. Newman (2006) shows the spectral algorithm for community detection as one of the most efficient methods. According to Newman (2006), “the problem of detecting and characterizing this community structure is one of the outstanding issues in the study of networked systems”.

The analysis of split graphs, as one of the common structures in terms of social networks’ modularity detection, represents the special interest for us.

2 Split graphs in social networks

The split graph is a graph that can be partitioned into the disjoint union of a clique and an independent set (Merris, 2003). In terms of social networks, split graphs reflect the realistic interpersonal and intergroup relations. It is very common to see social groups whose members are closely interrelated by similar ideas and interests, such as religion, research, education, level of income, etc. In terms of graph theory, these social groups form the structures called cliques (Luce and Perry, 1949). Finding the maximum clique, as the largest possible subgroup of closely related people in the social network, is an important step in the network’s analysis. It corresponds for searching of the most powerful group of people in the network. The problem of finding the maximum clique is NP-complete (Östergård, 2002). However, finding the maximum clique in small graphs is not problematic.

In contrast to cliques, independent sets in graphs, representing the sets of nodes with no edges connecting them (Boppana and Halldórsson, 1992), correspond to the socially fragmented and non-interrelated groups of people. Mostly, people in the independent sets do not know each other, have no common interests or may even compete with each other. Finding the maximum independent set is an NP-hard optimisation problem (Robson, 1986).

3 Concept of trust in social networks

Trust is the key concept in any social network. It is a basis for the formation of social groups and coalitions, identification of the most powerful nodes in the network, and it is the determinant for the information flow in social networks (Adali et al., 2010).

When a person has to decide whether to trust the other person or not, the decision about trust is influenced by a set of different factors. Following Adali et al. (2010), there are three basic decision components:

- 1 personal predisposition to trust
- 2 previous relationship with the person and his relatives, friends, colleagues, etc.
- 3 the opinion about the decisions and actions previously made by a third person.

Trust measuring in social networks is a complex process, and there are many trust models in the literature. For example, Abdul-Rahman and Hailes (2000) proposed the trust model in virtual communities based on the idea of measuring trust employing the mechanism of the experience and reputation of the network members. Considering trust as the 'subjective degree of belief about agents' (McKnight and Chervany, 1996; Misztal, 1996), Abdul-Rahman and Hailes (2000) show how to measure trust degrees and how to assign trust weights.

Another approach was invented by Aberer and Despotovic (2001). They present the algorithms of trust measuring based on the computation of the agent's reputation. The research shows a specific way of evaluating trustworthiness based on the local trust computations.

Adali et al. (2010) represent metrics of trust based on the analysis of dyadic relations. They described the idea of behavioural trust measures, which are based on determining the communication behaviour of agents in social networks. The represented methods of behavioural measures of trust were tested based on the Twitter network data.

Any trust relation is basically affiliated with the risk of making a wrong decision when communicating with the other persons in the social network. Buskens (2002) in the book *Social Networks and Trust* provides some examples about the risks of trust within the "social context" of trust relations.

Trust is highly associated with the risk of interconnection with wrong, dangerous or suspicious people and it is therefore important to assess the trustworthiness of the relations in social networks. According to Aberer and Despotovic (2001), "this allows to compute directly the expected outcome respectively risk involved in an interaction with an agent, and makes the level of trust directly dependent on the expected utility of the interaction". When discussing the problems of social exchange, Molm et al. (2000) specifically consider risk and trust as basic aspects in terms of classical exchange theory. According to that paper, the evaluation of trustworthiness is initially based on the analysis of risk and uncertainty of the exchange.

4 Problem description

In the given research, we are specifically interested in the analysis of social networks with the risk of trust as the basic social factor in the interpersonal communications and socio-economic exchange. We do not concentrate on the idea of how to measure the trust (or the risk of trust) in networks. This is the topic for different research, and we have described some approaches in this area in Section 3. Our goal is to analyse social networks in terms of weighted split graphs, where edge weights correspond to the risk of trust between nodes in a social network. Finding the maximum clique and one or more independent sets, we introduce the method for minimising the risk of interconnections between the maximum clique members and members of independent sets solving a k -cardinality assignment problem. In other words, we consider the clique as the socially powerful group of people, and we have independent sets of people who wish to enter the

clique (i.e., to become a member of the clique). In general, clique members do not wish persons from the ‘external world’ to enter their ‘internal environment’. From the clique’s position this is the risk that should be minimised or avoided completely. Nevertheless, complete avoidance is almost impossible, because social networks are not closed systems. In forming connections with the ‘external world’, clique members try to minimise their risk of interrelations. By applying semi-Lagrangian relaxation for the k -cardinality assignment problem, we minimise overall risks. Our mathematical formalisation of the semi-Lagrangian relaxation for the k -cardinality assignment problem is represented in the following Sections 5–6.

5 The integer programming formulation of the k -cardinality assignment problem

We construct the integer programming formulation of the k -cardinality assignment problem in the following way:

$$\text{Min } \sum_i \sum_j c_{ij} x_{ij} \quad (1)$$

Subject to:

$$\sum_{ij} x_{ij} = k \quad (2)$$

$$\sum_i x_{ij} \leq 1 \quad \forall j \in J \quad (3)$$

$$\sum_j x_{ij} \leq 1 \quad \forall i \in I \quad (4)$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in I, j \in J \quad (5)$$

6 Semi-Lagrangian relaxation

The formalisation of the semi-Lagrangian relaxation for the k -cardinality problem was formalised in Beltran et al. (2006).

We introduce our mathematical programming model of the semi-Lagrangian relaxation for the k -cardinality assignment problem:

$$\text{Max } SL(u) \text{ subject to } u \geq 0,$$

where $L(u)$ is defined by the following optimisation problem:

$$\text{Min } \sum_i \sum_j c_{ij} x_{ij} - u \left(\sum_{ij} x_{ij} - k \right) \quad (6)$$

Subject to:

$$\sum_{ij} x_{ij} \leq k \quad (7)$$

$$\sum_i x_{ij} \leq 1 \quad \forall j \in J \quad (8)$$

$$\sum_j x_{ij} \leq 1 \quad \forall i \in I \quad (9)$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in I, j \in J \quad (10)$$

The following theorem from Beltrán-Royo et al. (2012) gives the properties of the semi-Lagrangian dual theorem:

- 1 The semi-Lagrangian dual $L(u)$ is concave and $b - Ax(u)$ is a subgradient at u .
- 2 $L(u)$ is monotone and $L(u') \geq L(u)$ if $u' \geq u$ with strict inequality if $u' > u$ and $u \notin U^*$.
- 3 $U^* = U^* \cup \mathbb{R}_+^m$; thus U^* is unbounded.
- 4 If $x(u)$ is such that $Ax(u) = b$, then $u \in U^*$ and $x(u)$ solves the original problem.
- 5 Conversely, if $u \in \text{int}(U^*)$, then any minimiser $x(u)$ is optimal in the original problem.
- 6 The semi-Lagrangian relaxation closes the duality gap.

7 Implementation

We apply our mathematical programming model represented in Section 6 to analyse inter-component relations in split graphs in terms of the optimal links assignment. We introduce the method to calculate the overall optimal relations between the maximum clique (i.e., the most influential group of agents) and independent sets (i.e., disconnected group of agents that attempt to enter the clique) in a network:

```

TRUST-RISK_MINIMISATION G(V, E, W)
Input: graph G(V, E, W)
1 Maximum-clique-detection G(V, E, W)
2 Independent-sets-detection G(V, E, W)
3 Semi-Lagrangean relaxation for k links
Return: optimal set of links
    
```

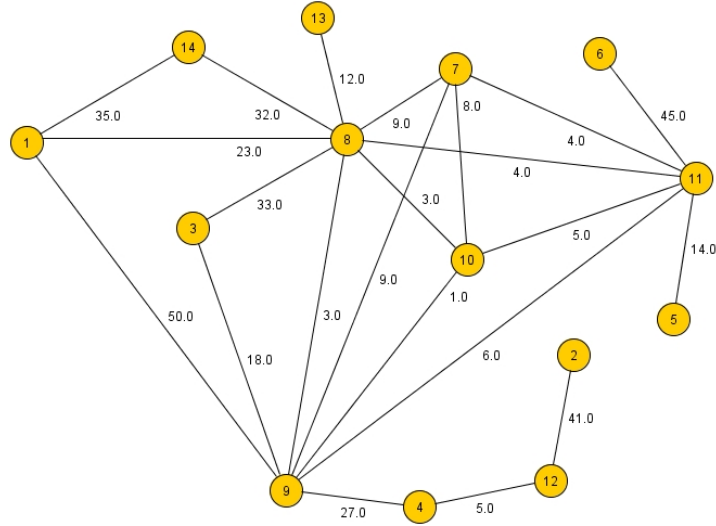
Notation: $G(V, E, W)$ is a graph with the set of nodes $G.V$, set of edges $G.E$, and the set of weights $G.W$; k is the number of links that has to be assigned between clique members and members of independent sets.

Algorithms for the maximum clique problem (i.e., line 1 in TRUST-RISK_MINIMISATION) are well-formalised in Carraghan and Pardalos (1990) and in Bomze et al. (1999). For independent sets detection (i.e., line 2) there are also many algorithms such as Bourgeois et al. (2010), Fomin et al. (2009) etc.

The mathematical programming model for k -cardinality assignment problem (i.e., line 3) is introduced in Section 6.

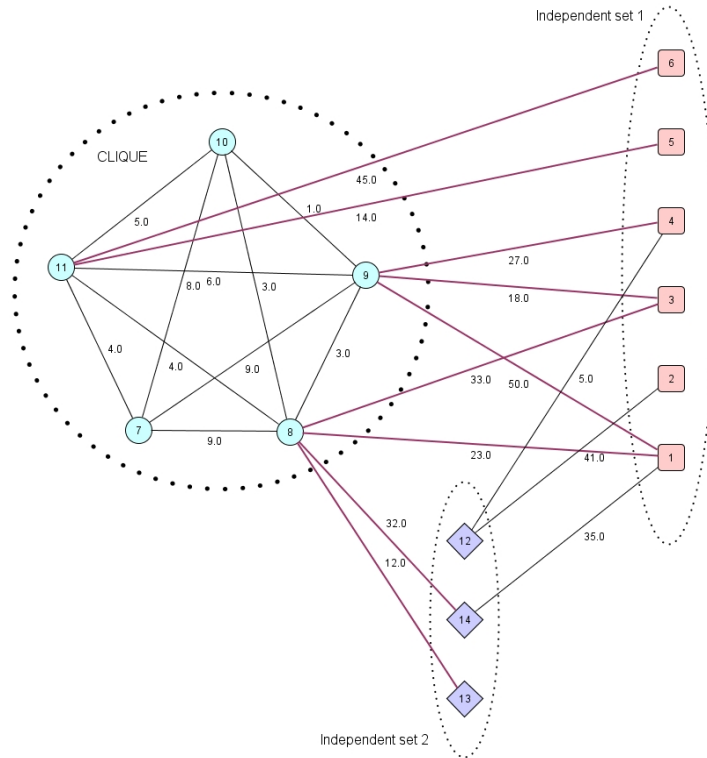
To show our approach we consider the network with 14 persons. The structure of the network is represented in Figure 1. The edges' weights correspond to the risk of trust between persons.

Figure 1 Tested social network (see online version for colours)



We apply TRUST-RISK_MINIMISATION and analyse the given graph as the split graph looking for the maximum clique and independent sets. The results of the analysis are represented in Figure 2.

Figure 2 Maximum clique and independent sets in the network (see online version for colours)



According to Figure 2, we have the clique that consists of 5 persons (i.e., 7, 8, 9, 10, and 11) and two independent sets: persons 1, 2, 3, 4, 5, 6 in the first independent set, and 12, 13, 14 – in the second independent set. Clique members have to communicate with external groups minimising the risk of being involved in interconnections with persons that are out of the clique.

In terms of social networks, there can be different situations. For example, clique is a research team of a university, where some research members are interested in hiring assistants for themselves from the set of students (i.e., independent set). Each clique member is interested in hiring the right person minimising the overall risk for the clique of hiring the wrong student from the independent set.

Another example is a company that has to choose specialists to hire from the number of trainees. In this case, the clique is the set of company departments looking for trainees to be hired. For instance, the IT department is looking for a programmer from one set of trainees (i.e., first independent set); the finance department might be looking for the financial analyst from another set of trainees (i.e., second independent set), etc.

Considering our split graph represented in Figure 2 we have $m = 5$, $n = 9$, $k = 3$, and the matrix of risks represented in Table 1.

Table 1 C-matrix of risks

		<i>Indep. set 1</i>						<i>Indep. set 2</i>		
		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>12</i>	<i>13</i>	<i>14</i>
CLIQUE	7	100	100	100	100	100	100	100	100	100
	8	23	100	33	100	100	100	100	12	32
	9	50	100	18	27	100	100	100	100	100
	10	100	100	100	100	100	100	100	100	100
	11	100	100	100	100	14	45	100	100	100

In C-matrix of risks m is the number of clique members, n is the number of persons in the independent sets, k is the number of persons to be involved in the interconnection with clique members, which we assign to be equal to ‘3’. Weights in the matrix correspond to the level of risk (i.e., risk of trust) with values in the range $[0, 100]$. It is important to notice that if an edge between the member of the clique and the member of the independent set does not exist then the risk of making the interconnection is assigned to be ‘100’ matching the highest level of risk and uncertainty.

We solve the k -cardinality assignment problem based on the semi-Lagrangean relaxation following the procedure described in Section 6.

Table 2 Cost matrix (see online version for colours)

		<i>Indep. set 1</i>						<i>Indep. set 2</i>		
		<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>12</i>	<i>13</i>	<i>14</i>
CLIQUE	7	X	X	X	X	X	X	X	X	X
	8	X	X	X	X	X	X	X	-6	X
	9	X	X	0	X	X	X	X	X	X
	10	X	X	X	X	X	X	X	X	X
	11	X	X	X	X	-4	X	X	X	X

Based on the optimisation results, the optimal semi-Lagrangean multiplier is $u^* = 18$. The resulting cost matrix is represented in Table 2.

In Table 2, all non-allowable assignments are marked by X. The solution to the problem is the ‘row-column’ assignment 8-13, 9-3, and 11-5 of cardinality ‘3’ with the objective function value equal to ‘-10’. The lower bound (LBD) is $54 - 10 = 44$. The feasible solution (the upper bound) is also ‘44’. Therefore, the optimal solution has been found.

The optimal semi-Lagrangean multiplier $u^* = 18$ has a meaningful interpretation in terms of social networks. It corresponds to the highest risk of interconnection with non-clique members: the interconnection between person 9 from the clique and person 3 from the first independent set is the most risky one. Two other newly assigned interconnections i.e., person 11 from the clique and person 5 from the first independent set; person 8 from the clique and person 13 from the second independent set, are less risky. In terms of social networks, it helps to formulate the prospective relation strategies with persons. For example, person 3 might be required to be controlled more than others or might be additionally trained. In general, solving the k -cardinality assignment problem, we minimise the risk of the clique members’ interconnections with new prospective members.

8 Interpretation of the results in terms of social networks

The social meaning of the described mechanism for the minimisation of risky intergroup interactions is important in terms of the trustworthiness analysis in social networks. The ability to detect the least risky relations between different types of social groups has a significant practical importance. Its central role is obvious in the research collaborations among scientists, joint ventures among companies, new employees hiring, and personnel management, etc.

The case, analysed in Section 7, represents a small social network with only 14 nodes. However, the real-world social networks are characterised by the large-scale structures and interrelations with complex nature. The analysis of risks in the interpersonal and intergroup communications is one important problem, but it is also important to know ways of minimising those risks.

9 Conclusions

In the given research, we considered social networks in terms of split graphs. Since split graphs reflect the structure of many real-world social networks, we analysed their internal mechanisms considering the risk of trust as the basic factor in the interpersonal and intergroup relations. In this paper we described the concept of trust explaining its exclusive importance. We considered the risk of trust between nodes in split graphs as a basis for links’ weights.

Next, we introduced the algorithm for the risk of trust minimisation in split graphs based on our approach of the semi-Lagrangean relaxation for the k -cardinality problem.

The overall algorithm consists of three steps:

- 1 max clique detection
- 2 independent sets detection
- 3 minimisation of risks in the inter-component relations in split graphs.

Steps 1 and 2 are well-formalised and described in literature. Step 3 is based on our approach for the semi-Lagrangian relaxation. It shows the effective mechanism to minimise the risky interconnections that has to be established in split graphs due to the internal social requirements.

References

- Abdul-Rahman, A. and Hailes, S. (2000) 'Supporting trust in virtual communities', *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2000*, January, IEEE, p.9.
- Aberer, K. and Despotovic, Z. (2001) 'Managing trust in a peer-2-peer information system', *Proceedings of the Tenth International Conference on Information and Knowledge Management*, October, ACM, pp.310–317.
- Adali, S., Escriva, R., Goldberg, M.K., Hayvanovych, M., Magdon-Ismael, M., Szymanski, B.K. and Williams, G. (2010) 'Measuring behavioral trust in social networks', *2010 IEEE International Conference on Intelligence and Security Informatics (ISI)*, May, IEEE, pp.150–152.
- Beltran, C., Tadonki, C. and Vial, J.P. (2006) 'Solving the p-median problem with a semi-Lagrangian relaxation', *Computational Optimization and Applications*, Vol. 35, No. 2, pp.239–260.
- Beltrán-Royo, C., Vial, J.P. and Alonso-Ayuso, A. (2012) 'Semi-Lagrangian relaxation applied to the uncapacitated facility location problem', *Computational Optimization and Applications*, Vol. 51, No. 1, pp.387–409.
- Bomze, I.M., Budinich, M., Pardalos, P.M. and Pelillo, M. (1999) 'The maximum clique problem', *Handbook of Combinatorial Optimization*, pp.1–74, Springer, USA.
- Boppana, R. and Halldórsson, M.M. (1992) 'Approximating maximum independent sets by excluding subgraphs', *BIT Numerical Mathematics*, Vol. 32, No. 2, pp.180–196.
- Bourgeois, N., Escoffier, B., Paschos, V.T. and van Rooij, J.M. (2010) 'A bottom-up method and fast algorithms for max independent set', *Algorithm Theory-SWAT 2010*, pp.62–73, Springer, Berlin, Heidelberg.
- Buskens, V. (2002) *Social Networks and Trust*, Series: Theory and Decision Library C, Vol. 30, Springer.
- Carraghan, R. and Pardalos, P.M. (1990) 'An exact algorithm for the maximum clique problem', *Operations Research Letters*, Vol. 9, No. 6, pp.375–382.
- Fomin, F.V., Grandoni, F. and Kratsch, D. (2009) 'A measure & conquer approach for the analysis of exact algorithms', *Journal of the ACM (JACM)*, Vol. 56, No. 5, p.25.
- Jackson, M.O. and Rogers, B.W. (2007) 'Meeting strangers and friends of friends: how random are social networks?', *The American Economic Review*, Vol. 97, No. 3, pp.890–915.
- Luce, R.D. and Perry, A.D. (1949) 'A method of matrix analysis of group structure', *Psychometrika*, Vol. 14, No. 2, pp.95–116.
- McKnight, D.H. and Chervany, N.L. (1996) *The Meanings of Trust*, Technical Report 94-04, Carlson School of Management, University of Minnesota.
- Merris, R. (2003) 'Split graphs', *European Journal of Combinatorics*, Vol. 24, No. 4, pp.413–430.
- Misztal, B.A. (1996) *Trust in Modern Societies: The Search for the Bases of Social Order*, Vol. 1, Polity Press, Cambridge.

- Molm, L.D., Takahashi, N. and Peterson, G. (2000) 'Risk and trust in social exchange: an experimental test of a classical proposition', *American Journal of Sociology*, Vol. 105, No. 5, pp.1396–1427.
- Newman, M.E. (2006) 'Modularity and community structure in networks', *Proceedings of the National Academy of Sciences*, Vol. 103, No. 23, pp.8577–8582.
- Östergård, P.R. (2002) 'A fast algorithm for the maximum clique problem', *Discrete Applied Mathematics*, Vol. 120, No. 1, pp.197–207.
- Robson, J.M. (1986) 'Algorithms for maximum independent sets', *Journal of Algorithms*, Vol. 7, No. 3, pp.425–440.