2012 International Conference on Economics, Business and Marketing Management
IPEDR vol.29 (2012) © (2012) IACSIT Press, Singapore

Optimal Selection of an Independent Set of Cliquesin a Market Graph

AthulaD. A. Gunawardena[1+], Robert R. Meyer[2], William L. Dougan[1],
Patrick E. Monaghan[1]and ChotonBasu[1]

[1] University of Wisconsin-Whitewater, Whitewater, WI 53190, USA.
[2]University of Wisconsin-Whitewater, Madison, WI 53706, USA.

Abstract. A complete market graph G on n markets is a simple undirected weighted complete graph in which vertices represent the markets and the weight associated with an edge is a correlation coefficient of the corresponding pair of markets calculated for a certain fixed period of time. A k-clique in G is an induced subgraph of G having k vertices.Given a market graph, we construct a clique graph whose vertices represent the k-cliques in the market graph, and define edges and weights based on certain correlation functions defined on pairs of cliques. An independent set of cliquesconsists of members that have negative clique correlations with respect to each other.Our goal is to find an optimal set of independent cliques in a clique graph subject to certain constraints in order to produce highly correlated cliques which are pairwise anticorrelated based on clique correlations.Such a solution provides useful information about marketbehaviours and can be utilized for portfolio optimization.We present integer linear and quadratic program (IP) formulations to find optimal solutions for the above problem.

Keywords: market graph, clique graph, integer program.

1. Introduction

Mutual relations of members in a set of markets (or financial instruments) can be naturally represented by a simple undirected weighted graph called a *market graph*. In a market graph, each financialinstrument is represented by a vertex, and two vertices are connected by anedge if the correlation coefficient of the corresponding pair of instruments (calculated for a certain fixed period of time) exceeds a specified threshold. A *complete market graph* is a market graph constructed without applying a threshhold (contains all the edges).More details of market graphs can be found in [1], [2], [3], [4].

A *clique* in an undirected graph is a subset of its vertices such that every two vertices in the subset are connected by an edge.This is equivalent to saying that the subgraph induced by the subset of vertices is complete.For a positive integer *k*, a *k-clique* is a clique having *k* vertices.An *independent set* is a set of vertices in a graph, no two of which are adjacent.For a positive threshold value, aclique in a market graph will represent a set of instruments whose prices change similarly over time (a change of the price of any instrument in a clique is likely to affect all other instruments in this clique). For a negative threshold value, an independent setwill consist of instruments that have negative correlations with respect toeach other, therefore, they can be treated as a so-called diversified portfolio.Analyzingcliques and independent sets in a market graph gives us a very valuableknowledge about the internal structure of the given set of markets, and various studies about cliques and independent sets can be found in [5], [7], [6].

In this paper, we extend the above framework to study the group behavior of markets by constructing a new graph called a *clique graph*.Given a complete market graph *G* and a positive integer *k*, we construct a clique graph, with all the *k*-cliques of *G* as its vertices, and by defining edges according to a threshold on a correlation function defined on pairs of *k*-cliques.Our goal is to find an optimal set of independent cliques in a clique graph that satisfy certain constraints to produce highly correlated cliques which are pairwise

+ *E-mail address*: gunawara@uww.edu

anticorrelated based on clique correlations. We present integer linear and quadratic program (IP) formulations to find optimal solutions for the above problem.

2. Preliminaries

Let M = $\{M_1, M_2, \ldots, M_n\}$ be a set of *n* markets and *C* be an *n* × *n* real matrix where the *(i,j)*th entry of *C* (c_{ij})represents a correlation coefficient between M_i and $M_j$calculated for a certain fixed period of time. Hence -1 ≤ c_{ij} ≤1 for 1 ≤ *i,j* ≤ *n*. Let Ebe $\{ e_{ij} = (M_i, M_j) \mid 1 \le i,j \le n \}$.We construct a complete market graph *G*= (M, E) with Mas its set of vertices, E as its set of edges, and the weight associated with the edge e_{ij}as c_{ij}. For a given positive integer *k (k < n),* letQ^k be the set of *k*-cliques in *G*. Since *G* is complete, Q^k=$\{Q \mid Q \subseteq M, |Q| = k\}$. We define a weight function *w*: , Q^k→[0,1] such that for any *Q*∈ Q^k,

$$\mathrm{w(Q)} = \frac{1}{k^2} \sum_{M_i, M_j \in Q} c_{ij}.$$

We define a correlation function *f*: $Q^k \times Q^k$→ [-1,1] such that for any $(Q_1, Q_2) \in Q^k \times Q^k$,

$$\mathrm{f}(Q_1, Q_2) = \frac{1}{k^2} \sum_{M_i \in Q_1, M_j \in Q_2} c_{ij}.$$

A clique is called a *positively correlated (or anticorrelated)*clique if it has a positive (or negative) weight.A pair of cliques is called a *positively correlated (or anticorrelated)* pair if it has a positive (or negative) correlation function *(f)* value.Two cliques Q_1 and $Q_2$are *disjoint*if they do not have common members.

For any β∈ [-1,1], we define a clique graph *CG*as follows. The set of vertices of *CG* is Q^k. For $Q_1, Q_2 \in Q^k$,$Q_1 \neq Q_2$, there is an edge between Q_1andQ_2if $\mathrm{f}(Q_1, Q_2) > \beta$.Note that if β is negative,any independent set in *CG* is pairwise anticorrelated.For our framework, we define an *independent set of cliques* in *G* as an independent set of disjoint cliques in a clique graph with β = 0.Thus the members of an independent set of cliques in *G* are disjoint and not pairwise positively correlated in *CG*.For a diverse market graph, our goal is to find a set of independent highly correlated cliques.

2.1. Independent Clique Set (ICS) Problem

Let *G*be a complete market graph on *n*markets and *CG* be a derived clique graph of Gwith its *k*-cliques and β= 0. Let *N*,Δbe positive integers such that *kN* ≤ *n*, and Δ ≤ 1. Let $\{\lambda_i \mid 1 \le i \le N\}$and $\{ \gamma_{ij} \mid 1 \le i,j \le N, i \neq j \}$ be sets of positive real numbers.The ICS problem is stated as follows.

$$\text{Maximize } [\sum_{i=1}^{N} \lambda_i w(Q_i) - \sum_{1 \le i,j \le N, i \neq j} \gamma_{ij}\, f(Q_i, Q_j)]$$

subject to

$$Q_i \in Q^k \text{ for } 1 \le i \le N,$$
$$Q_i \cap Q_j = \emptyset \text{ for } 1 \le i,j \le N, i \neq j,$$
$$w(Q_i) \ge \Delta \text{ for } 1 \le i \le N,$$
$$\mathrm{f}(Q_i, Q_j) \le 0 \text{ for } 1 \le i,j \le N, i \neq j.$$

Note that the ICS problem is always feasible if we remove threshold constraints $w(Q_i) \ge \Delta$ and $\mathrm{f}(Q_i, Q_j) \le 0$. Since the objective function has competing components, we can select the parameters λ_i, $\gamma_{ij}$and Δ to generate solutions with various characteristics.

3. Integer Programming Models

In this section, we present linear and quadratic formulations of integer programming models that are solved to generate optimal solutions for the ICS problem.Let *G* be a complete market graph on *n* markets and *C* be its associated*n* × *n* correlation matrix. Let CG, k, N,λ_i, $\gamma_{ij}$and Δ as mentioned above.

3.1. Integer Linear Program Model

$$\text{Maximize } [\sum_{i=1}^{N} \lambda_i w(Q_i) - \sum_{1 \le i,j \le N, i \ne j} \gamma_{ij} f(Q_i, Q_j)]$$

subject to

$$w(Q_r) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} u_{ij}^r \text{ for } 1 \le r \le N,$$

$$w(Q_r) \ge \Delta \text{ for } 1 \le i \le N,$$

$$f(Q_s, Q_t) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} v_{ij}^{st} \text{ for } 1 \le s,t \le N, s \ne t,$$

$$\mathrm{f}(Q_s, Q_t) \le 0 \text{ for } 1 \le s,t \le N, s \ne t,$$

$$\sum_{i=1}^{n} x_i^r = k \quad \text{for } 1 \le r \le N,$$

$$\sum_{i=1}^{n} x_i^r = 1 \quad \text{for } 1 \le i \le n,$$

$$u_{ij}^r \le x_i^r, \qquad u_{ij}^r \le x_j^r,$$

$$u_{ij}^r \ge x_i^r + x_j^r - 1,$$

$$v_{ij}^{st} \le x_i^s, \qquad v_{ij}^{st} \le x_j^t,$$

$$v_{ij}^{st} \ge x_i^s + x_j^t - 1.$$

where $x_i^r, u_{ij}^r, v_{ij}^{st}$ are binary variables for $1 \le r, s, t \le N$ and $1 \le i,j \le n$. We have implemented this integer linear model in GAMS/CPLEX and found it to be effective for small values of n(i.e., $n \le 30$). The following integer quadratic program model provides optimal solutions for the ICS problem for $n \approx 500$ in less than a minute.

3.2. Integer Quadratic Program Model

$$\text{Maximize } [\sum_{i=1}^{N} \lambda_i w(Q_i) - \sum_{1 \le i,j \le N, i \ne j} \gamma_{ij} f(Q_i, Q_j)]$$

subject to

$$w(Q_r) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_i^r x_j^r \text{for } 1 \le r \le N,$$

$$w(Q_r) \ge \Delta \text{ for } 1 \le i \le N,$$

$$f(Q_s, Q_t) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_i^s x_j^t \text{for } 1 \le s,t \le N, s \ne t,$$

$$\mathrm{f}(Q_s, Q_t) \le 0 \text{ for } 1 \le s,t \le N, s \ne t,$$

$$\sum_{i=1}^{n} x_i^r = k \quad \text{for } 1 \le r \le N,$$

$$\sum_{i=1}^{n} x_i^r = 1 \quad \text{for } 1 \le i \le n,$$

where $x_i^r, 1 \le i \le n, 1 \le r \le N,$ are binary variables.

4. Results

We have selected diverse collection of 72 commodity markets and indexes shown in Table 1.

Table 1: Selected Commodity Markets and Indexes

Market or Index	Market or Index
1. Australian Dollar (AD,CME)	37. Russell 2000 Index (RL, CME)
2. British Pound (BP, CME)	38. Russell 2000 Index E-Mini (ER, CME)
3. Canadian Dollar (CD, CME)	39. S&P 500 (SP, CME)
4. EuroFX (EC, CME)	40. S&P E-Mini (ES, CME)
5. Japanese Yen (JY, CME)	41. S&P Midcap 400 (MD, CME)
6. Mexican Peso (MP, CME)	42. Treasury Notes 10 Year (TY,CBOT)
7. New Zealand Dollar(NE,CME)	43. Treasury Notes 2 Year (TU, CBOT)
8. Swiss Franc (SF, CME)	44. Treasury Notes 5 Year (FV, CBOT)
9. Heating Oil (HO, NYMEX)	45. U.S. Dollar Index (DX, NYBOT)
10. Light Crude Oil (CL,NYMEX)	46. Value Line Index (MV, KCBT)
11. Light Crude Oil EmiNY (QM, NYMEX)	47. Western Barley (AB, WCE)
12. Natural Gas (NG, NYMEX)	48. Canola (RS, WCE)
13. PJM Western Electricity (JM, NYMEX)	49. Corn (C, CBOT)
14. Propane (PN, NYMEX)	50. Corn Mini (YC, CBOT)
15. Unleaded Gas (HU, NYMEX)	51. Cotton (CT, NYBOT)
16. BFP Milk (DA, CME)	52. Feed Wheat (WW, WCE)
17. Butter (DB, CME)	53. Hard Red Spring Wheat (MW, MGE)
18. Cocoa (CC, NYBOT)	54. Kansas Wheat (KW, KCBT)
19. Coffee (KC, NYBOT)	55. Oats (O, CBOT)
20. Lumber (LB, CME)	56. Rice (RR, CBOT)
21. Sugar #11 (SB, NYBOT)	57. Soybean Meal (SM, CBOT)
22. Sugar #14 (SE, NYBOT)	58. Soybean Oil (BO, CBOT)
23. 30 Year US Treasury Bonds (US, CBOT)	59. Soybeans (S, CBOT)
24. CRB Index Futures (CR, NYBOT)	60. Soybeans Mini (YK, CBOT)
25. Dow Jones Industrial Av. Futures (DJ, CBOT)	61. Wheat (W, CBOT)
	62. Wheat Mini (YW, CBOT)
26. Euro Dollar (ED, CME)	63. Cattle Feeder (FC, CME)
27. Euro Yen (EY, CME)	64. Cattle Live (LC, CME)
28. Federal Funds 30 Day (FF, CBOT)	65. Hogs Lean (LH, CME)
29. Goldman Sachs Index (GI, CME)	66. Pork Bellies (PB, CME)
30. LIBOR - 1 Month (EM, CME)	67. Aluminum (AL, COMEX)
31. Mini Dow Jones (YM, CBOT)	68. Copper High Grade (HG, COMEX)
32. Municipal Bonds (MB, CBOT)	69. Gold 100 oz. (GC, COMEX)
33. NASDAQ 100 E-mini Futures (NQ, CME)	70. Palladium (PA, NYMEX)
34. NASDAQ 100 Index Futures (ND, CME)	71. Platinum (PL, NYMEX)
35. Nikkei Index (NK, CME)	72. Silver 5000 oz. (SI, COMEX)
36. NYSE Composite Index (YX, NYBOT)	

We used the day to day percentage variation of closing prices to calculate correlationmatrices for a 500 day period with different starting dates. We use our quadratic model to find optimal solutions to select an independent set containing three 5-cliques.Table 2 shows clique correlations of an optimal solution containing three 5-cliques (Q_1, Q_2, Q_3) for a 500 day period starting from January 01, 2007. The diagonal entries are clique weights.

Table 2: Clique Correlations of an optimal solution containing three 5-cliques (Q_1, Q_2, Q_3) for a 500 day period starting from January 01, 2007

	Q_1	Q_2	Q_3
Q_1	0.995	-0.459	-0.553
Q_2	-0.459	0.999	-0.565
Q_3	-0.533	-0.565	0.944

Table 3 shows the individual members of Q_1, Q_2, and Q_3.Note that within each of these cliques the members are highly correlated, and that diversity between cliques is also high.We used the weighting parameters $\lambda_1 = \lambda_2 = \lambda_3 = 0.5$ and $\gamma_{ij} = 1.0$ for $1 \leq i,j \leq 3$, $i \neq j$.

Table 3: The members of the three 5-cliques (Q_1, Q_2, Q_3)

Clique	Market or Index	Type
Q_1	Propane (PN, NYMEX) BFP Milk (DA, CME) Treasury Notes 10 Year (TY, CBOT) Treasury Notes 2 Year (TU, CBOT) Treasury Notes 5 Year (FV, CBOT)	Energy Exotic Financial Index Financial Index Financial Index
Q_2	Canadian Dollar (CD, CME) EuroFX (EC, CME) Swiss Franc (SF, CME) Euro Dollar (ED, CME) Platinum (PL, NYMEX)	Currencies Currencies Currencies Financial Index Metals
Q_3	Mini Dow Jones (YM, CBOT) NASDAQ 100 E-mini Futures (NQ, CME) S &P E-Mini (ES, CME) Western Barley (AB, WCE) Canola (RS, WCE)	Financial Index Financial Index Financial Index Grains Grains

5. Conclusion

In this paper, we constructed a clique graph and used its properties to study group behaviors of markets. The above results show that our quadratic program model is very effective in choosing optimal solutions for the ICS problem. These integer programming models can be extended to find groups of markets with different characteristics. We plan to extend these results to find themaximal independent sets in clique graphs over different correlation functions.

6. Acknowledgements

The authors thank Professor Leon Arriola forhis constructive and helpful comments.

7. References

[1] V. Boginski, S.Butenko, and P. M. Pardalos, On structural properties of the market graph, A. Nagurney, editor, *Innovation in Financial and Economic Networks*, Edward Elgar Publishers, London. 2003, pp. 29--45.

[2] V. Boginski, S.Butenko, and P. M. Pardalos, Modeling and optimization in massive graphs, P. M. Pardalos and H. Wolkowicz, editors, *Novel Approaches to Hard Discrete Optimization*, American Mathematical Society. 2003, pp. 17--39.

[3] V. Boginski, S.Butenko, and P. M. Pardalos,Network models of massive datasets, *Computer Science and Information Systems*. 2004, **1**:79--93.

[4] V. Boginski, S.Butenko, and P. M. Pardalos, Mining market data: a network approach, *Computers & Operations Research*. 2006, **33**:3171-3184.

[5] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness*, W.H. Freeman and Company, New York, 1979.

[6] P. M. Pardalos and J. Xue, The maximum clique problem, *Journal of Global Optimization*. 1994, **4** (3): 301—328.

[7] C. Q. Zhang, Finding critical independent sets and critical vertex subsets are polynomial problems, *SIAM J. Discrete Mathematics*. 1990, **3**: 431—438.