InvestorCliques (796315) – EU-Project Milestone 2.1

Clique Problem*

Le Viet Hung†

Department of Computing Sciences,

Tampere University, Tampere Finland

* I thank Mr. Kestutis Baltakys and Prof Juho Kanniainen for comments and support. A joint paper on the same topic will be published later.

† http://www.investorcliques.eu.; hung.le@tuni.fi

SUMMARY

Many real world problem of complex systems are formalized as maximal clique enumeration or maximum clique detection. For example, in the protein structure similarity network, clique represents for the structure similarity between known and unknown function proteins. In the protein protein interaction network, cliques are protein complexes which signify a biological function. In the collaboration network, large cliques are formed when project involves many participants. In the World Wide Web, cliques are set of pages which are fully connected.

In our context, investor stock trading network, cliques are investors behave similarly in terms of transactions and they might have similar portfolios with a given distance measure. The cliques could represent lead investors who change trading directions first and who are followed by others in a herd, aggressive investors that have volatile trading strategies, or passive investors who rarely rebalance their portfolios.

Clique problem is the computational problem of finding cliques (complete subgraphs) in a graph. The problem includes finding a maximum clique (a clique with the largest number of nodes) and enumerating all maximal cliques. The purpose of this milestone is to investigate a number of recently reported exact algorithms for the clique problem and their capacity on investor network.

Maximum clique detection

We implemented the exact algorithms for solving the problem of maximum clique detection: the central clique solver of Carraghan and Pardalos [1] and the one of Pattabiraman *et al.* [2], Rossi *et al.* [3], Verma *et al.* [4] for massive sparse graph. In addition, we incorporate novel pruning techniques based on the investor behavior context. For instance, to detect a large herding pattern when the bubble propagates, we might expect a large clique in the network by setting a high input value of minimum size. Another example, we can limit the search space to the specific branches which contain household investors or present overexpression of specific investor attributes. A more detailed technical Python implementation of the algorithm is presented at: http://www.investorcliques.eu/category/programming/

Maximal clique enumeration

The algorithm of Eppstein and Strash [5] is self implemented. We use networkx package[6] to implement the one of Tomita *et al.* [7]. Similar to maximum clique detection, we incor-

porate novel pruning techniques based on the investor behavior context.

DATA

We first use the input correlation network from our lab's previous work [8]. The data used in this study is the central register of shareholdings for Finnish stocks from Finnish central depository, provided by Euroclear Finland. Our sample data consists of the market-place transactions of 100 Finnish stocks consisting of investor's transactions around dot-com bubble from 1 January 1998 to 1 January 2002. A more detailed description of the data set is provided in Refs [8–12].

AVAILABILITY

Source code Python can be found at: http://www.investorcliques.eu/category/programming/

REFERENCE

https://gist.github.com/abhin4v/8304062 https://networkx.github.io/documentation/latest/$_m$*odules/r*

[1] R. Carraghan and P. M. Pardalos, Operations Research Letters **9**, 375 (1990).

[2] B. Pattabiraman, M. M. A. Patwary, A. H. Gebremedhin, W.-k. Liao, and A. Choudhary, in *International Workshop on Algorithms and Models for the Web-Graph* (Springer, 2013) pp. 156–169.

[3] R. A. Rossi, D. F. Gleich, A. H. Gebremedhin, M. Patwary, and M. Ali, arXiv preprint arXiv:1302.6256 (2013).

[4] A. Verma, A. Buchanan, and S. Butenko, INFORMS Journal on computing **27**, 164 (2015).

[5] D. Eppstein and D. Strash, in *International Symposium on Experimental Algorithms* (Springer, 2011) pp. 364–375.

[6] A. Hagberg, P. Swart, and D. S Chult, *Exploring network structure, dynamics, and function using NetworkX*, Tech. Rep. (Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008).

[7] E. Tomita, A. Tanaka, and H. Takahashi, Theoretical Computer Science **363**, 28 (2006).

[8] S. Ranganathan, M. Kivelä, and J. Kanniainen, PloS one **13**, e0198807 (2018).

[9] M. Grinblatt and M. Keloharju, Journal of financial economics **55**, 43 (2000).

[10] M. Tumminello, F. Lillo, J. Piilo, and R. N. Mantegna, New Journal of Physics **14**, 013041 (2012).

[11] K. Baltakys, J. Kanniainen, and F. Emmert-Streib, Scientific reports **8**, 8198 (2018).

[12] F. Musciotto, L. Marotta, J. Piilo, and R. N. Mantegna, Palgrave Communications **4**, 92 (2018).