InvestorCliques (796315) – EU-Project Milestone 2.2

Clique percolation*

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

Community structure of the complex network provides the useful information and reveals the function of the system. Many community detection methods developed so far have shown a good performance on the existing benchmark networks. However how good they are for the investor network, in terms of both accuracy and computing time, remains still open. Among class of community detection algorithms, clique percolation has shown a good accuracy and capability of detecting the overlapping and nested community. This is the case of most of the actual systems which have highly overlapping cohesive groups of nodes. The purpose of this milestone is to implement the existing clique percolation methods and investigate the overlapping community of investor network.

Clique percolation

First of all, the investor stock trading network is edge weighted network. The input network data is retrieved from [1] and filtered out (using milestone 1.2) to keep only links which have significant p-value or positive weight. Then all maximal cliques are enumerated using the algorithm of Eppstein and Strash [2] implemented in the milestone 2.1 of this project.

Clique graph is created with nodes are cliques and links are the overlap between them. Finally, the connected components of clique graph spell out the communities. The routine of percolation method is implemented in networkx package[3].

We implemented the original algorithm of Palla *et al.* [4]. It is applied only for binary network. To apply this method, we need to transform the weighted network into the binary one by keeping only links which have weight larger than input threshold. The threshold value of the weight and the k are optimal if the resulting community structure are as diverse as possible[5]. More specifically, Palla *et al.* [4], Farkas *et al.* [6] suggest that the threshold is optimal if the largest community is twice the size of the second largest one.

Furthermore, we implemented the algorithm of Farkas *et al.* [6] which can be used for weighted network and are slightly different from the original one. Specifically, they introduce clique intensity and consider only cliques which have size larger than k and intensity larger than critical input value. Those optimal input values are well documented at the work of Kumpula *et al.* [5], Farkas *et al.* [6]. Thus the algorithm allows both strong and weak links to be in the cliques as long as their geometric mean (clique intensity) larger than the critical value.

A more detailed technical Python implementation of the algorithm is presented at: http://www.investorcliques.eu/category/programming/

DATA

We first use the input correlation network from our lab's previous work [1]. 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 marketplace 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 [1, 7–10].

AVAILABILITY

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

REFERENCE

https://gist.github.com/abhin4v/8304062

https://gmarti.gitlab.io/networks/2018/06/03/pmfg-algorithm.html

https://github.com/hagberg/planarity

http://jgaa.info/accepted/2004/BoyerMyrvold2004.8.3.pdf

https://networkx.github.io/documentation/stable/index.html

---

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

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

[3] 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).

[4] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature **435**, 814 (2005).

[5] J. M. Kumpula, M. Kivelä, K. Kaski, and J. Saramäki, Physical Review E **78**, 026109 (2008).

[6] I. Farkas, D. Ábel, G. Palla, and T. Vicsek, New Journal of Physics **9**, 180 (2007).

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

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

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

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