Community Detection in A Network of Money Circulation

Juan Estrada[*]   Diego Rojas[†]   Yinghui Dong[‡]

April 14, 2019

1   Introduction

Network structures shape the observed outcomes across different types of markets. Human behavior like crime, substances abuse, educational achievement among others is significantly affected by social interactions (Jackson, 2008). More complex structures developed by individuals including financial systems are also governed by network interactions. The importance of networks in determining market outcomes is translated into an empirical relevance of understanding what characteristics of the network are relevant. In this regard, community detection has become a key task from both the social and computational perspectives (Jackson and Evan, 2019; Du et al., 2017b).

One particular market where the network structure is a cause and a consequence of human social interaction is the market of money circulation in a country. If we think about the economic structure of a nation as a human body, bank bills are comparable with the red blood cells in the human circulatory system. This analogy leads to an important implication: understanding bills' individual behavior will enlighten the mechanisms determining the operation of the complete economic system. This article is divided into two main components. The initial part of the paper presents the implementation of the algorithm developed by Du et al. (2017a). In this section, we provide a test for the performance of the algorithm using a simulated data set. Additionally, we show differences in complexity when changing the way how the minimization problems are calculated. In the second part of the article, we use the developed algorithm to perform a clustering procedure to the network of money circulation in Canada. This method allows us to find existing communities in the network while also using important observable characteristics of the bank notes (bank notes and notes are going to be used as interchangeably).

In particular, this document uses the Information Management System Program study of bank notes from the Bank of Canada (BoC). This is a unique data set containing records that track the

---

[*]Department of Economics, Emory University, Ph.D. Student in economics. E-mail: jjestra@emory.edu.

[†]Department of Economics, Emory University, Ph.D. Student in economics. E-mail: drojasb@emory.edu.

[‡]Department of Computer Science, Georgia Institute of Technology, Ms.c. Student in computer science.

events in the "life" of every note issued by the Bank of Canada. For this document, we have a sample of around 300 million note scan records. The period for which data is obtained is the year from August 2017 to July 2018. We can build networks at two levels: the region level and the financial institutions level networks. Specifically, the financial institution that receives the note from the Bank, and the financial institution that deposits back the note are known.

The most attractive feature of this analysis is that the richness of the data allows to observe the circulation patterns of every single note existing in the Canadian economy. The paper focuses on the network spanned by the circulation of the notes across regions and financial institutions to perform the community detection excessive. With the available information, in addition to finding relevant communities, other type of questions related with the velocity of money, heterogeneity in the circulation patters of different bills' denominations, and regional differences in the quantity and frequency of the received notes can be answered.

2   Related work

This article can relate with two different literatures. From the computational perspective, this article belongs to the literature of community detection methods. A well structured summary of the development of methods of community detection can be found in Du et al. (2017b). The authors present a complete review regarding joint matrix factorization. The method implemented in this paper has multiple advantages over previous approaches. It improves how interpretable are the matrix resulting from the decomposition. Additionally, it overcomes some computationally difficulties based on the non-negative constrain of all factors. In the same vein, there have been developments regarding clustering methods, 201 (2018) present a relevant work with the advantage of including clustering of directed graphs.

From the perspective of community detection in backing, the literature is not developed yet. Theoretically, the detection of existent communities is relevant because of the possibility of contagion of idiosyncratic shocks through the network structure in the financial system (Gai and Kapadia, 2010). Related with the existence of communities is the density of the connection for some particular subgraphs of the total network and it may also enhances financial stability (Acemoglu et al., 2015). Despite the theoretical contributions regarding the importance of communities to predict financial stability, to our knowledge, there are empirical papers developing banking community detection. Our paper contributes to fill the gap of empirical evidence of the existence of financial communities in the network of money circulation.

3   The problem and context

As mentioned, the goal of the paper is to efficiently perform a community detection of a network dataset containing additional information of the nodes. To achieve the objective, we implement the hybrid clustered algorithm developed by Du et al. (2017b). In our case, we have information of the principal financial institutions in Canada. For those institutions we can observe an identifier of the bank receiving transfers of bank notes from the Bank of Canada. Analogously, we observe the financial institution that deposits back the bank notes. With this information, we define a connection between two financial institutions A and B as a situation where *Bank A* receives the note x from the Bank of Canada which is followed by *Bank B* depositing the same note back to the Bank of Canada. Each connection between two financial institutions varies in the amount of bank notes that are received and deposited. This yields a network of size 50. Note that this definition generates a directed network because the number of notes sent from Bank A to B may be different from the number sent from Bank B to A.

The algorithm in Du et al. (2017b) is designed for symmetric adjacency matrices. To create a symmetric matrix, we redefine the connections as the total amount of transactions between banks A and B. For each bank note, we have a total of 22 fit characteristics, 5 denominations and we can observe whether the bill was transfer or deposited to the BoC. Using the notes' characteristics, we create aggregate features for each financial institution. To capture the possible heterogeneity in bank notes' fit, we use the 10 deciles of the empirical distribution of fit characteristics within each financial institution instead of just utilizing sample means. This procedures give us a total of 2,200 observable characteristics for each bank. Therefore, our matrix of observable information is defined by $X \in \mathbb{R}_+^{m \times n}$ and the graph structure is represented the symmetric matrix $S \in \mathbb{R}_+^{n \times n}$ where $m = 2,200$ and $n = 50$. The objective is to find a low rank representation that simultaneously account for the network structure and the characteristics matrix information. The objective function formalizing this idea is given by:

$$\min_{W \geq 0, H \geq 0} ||X - WH||_F^2 + \alpha ||S - H^T H||_F^2, \tag{3.1}$$

where α is the weighting parameter, $W \in \mathbb{R}_+^{m \times k}$ and $H \in \mathbb{R}_+^{k \times n}$ where k is much smaller than m and n. The article propose a block coordinate descent (BCD) scheme to find the solution for the objective function which require to solve the following three subproblems:

$$\min_{W \geq 0} ||H^t W^t - X^t||_F \tag{3.2}$$

$$\min_{\tilde{H} \geq 0} \left\| \begin{bmatrix} \sqrt{\alpha}H^t \\ \sqrt{\beta}I_k \end{bmatrix} \tilde{H} - \begin{bmatrix} \sqrt{\alpha}S \\ \sqrt{\beta}H \end{bmatrix} \right\|_F \tag{3.3}$$

$$\min_{\tilde{H} \geq 0} \left\| \begin{bmatrix} W \\ \sqrt{\alpha}\tilde{H}^t \\ \sqrt{\beta}I_k \end{bmatrix} H - \begin{bmatrix} X \\ \sqrt{\alpha}S \\ \sqrt{\beta}\tilde{H} \end{bmatrix} \right\|_F . \tag{3.4}$$

$$\tag{3.5}$$

Note that each subproblem is a nonnegative least squares (NLS). The procedure is based on an iterative algorithm that has been proved to converge. The solution of the NLS can be based on different basic decompositions. In the next sections, we show that the algorithm that implement works for a testing data set, we change the basic decomposition and show how the performance of the algorithm change, and finally we apply it to our empirical dataset.

References

2018. Codinmf: Co-clustering of directed graphs via nmf. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018* : 3611–3618.

Acemoglu D, Ozdaglar A, Tahbaz-Salehi A. 2015. Systemic risk and stability in financial networks. *The American Economic Review* **105**: 564–608. ISSN 00028282. URL http://www.jstor.org/stable/43495393

Du R, Drake B, Park H. 2017a. Hybrid clustering based on content and connection structure using joint nonnegative matrix factorization. *Journal of global optimization* : 1–17ISSN 0925-5001.

Du R, Kuang D, Drake B, Park H. 2017b. Hierarchical community detection via rank-2 symmetric nonnegative matrix factorization. *Computational Social Networks* **4**: 7. ISSN 2197-4314. URL https://doi.org/10.1186/s40649-017-0043-5

Gai P, Kapadia S. 2010. Contagion in financial networks. *Proceedings: Mathematical, Physical and Engineering Sciences* **466**: 2401–2423. ISSN 13645021. URL http://www.jstor.org/stable/25706352

Jackson MO. 2008. *Social and economic networks.* ISBN 9780691134406.

Jackson MO, Evan S. 2019. Behavioral Communities and the Atomic Structure of Networks. *SSRN* .