

Visualization of a directed network with focus on its hierarchy and circularity^{*}

Yuichi Kichikawa¹, Takashi Iino¹, Hiroshi Iyetomi¹, and Hiroyasu Inoue²

¹ Niigata University, Niigata, Japan

ykichikawa@m.sc.niigata-u.ac.jp

² The University of Hyogo, Kobe, Japan

Abstract. The spring-electric model is a useful tool to visualize a large-scale complex network. However, information on the flow of directed network may not be properly reflected because links are basically treated as undirected. Here, we propose a new visualization method with an explicit account of network flow structure information by combining Helmholtz-Hodge decomposition and the spring-electric model. We then demonstrate its effectiveness by adopting actual Japanese production flow network as a test ground. The Helmholtz-Hodge decomposition enables us to break down flow on a directed network into two flow components: potential flow and circular flow. The potential flow between a pair of nodes is given by difference of their potentials, and hence, the potential of a node shows its hierarchical position in a network. On the other hand, the circular flow component illuminates feedback loops built in a network. We also identify dominant clusters of firms forming feedback loops by applying a flow-based community detection method to the extracted circular flow network. We find that both hierarchical and loop structures coexist within the major industries such as construction, manufacturing, and wholesales.

Keywords: Visualization · directed graph · Helmholtz-Hodge decomposition · community detection · production network.

1 Introduction

It is difficult to analyze all connections in complex networks, because of their multiplicity and complexity. Visualization, which is gaining popularity owing to the recent development of graphics technology, is a useful tool with which to illuminate structural properties of networks. Appropriate depiction of a complex network greatly helps in grasping its intricate structures by providing an intuitive understanding. Various algorithms have been developed to visualize networks. A

^{*} This research was supported by MEXT as Exploratory Challenges on Post-K computer (Studies of Multi-level Spatiotemporal Simulation of Socioeconomic Phenomena, Macroeconomic Simulations). We used computational resources of the K computer provided by the RIKEN Center for Computational Science through the HPCI System Research project (Project ID:hp170242, hp180177)

spring-electric model is adopted in which pairs of nodes with direct relations are physically connected by springs, and nodes in any pair repel each other through a repulsive Coulomb force [6]. The attractive force of the spring keeps intimate nodes close in space. On the other hand, the repulsive Coulomb force tends to distribute firms uniformly over the available space and prevents entanglement of the network. Although the spring-electric model is a useful method of visualization, information on the flow of the network can't be reflected since the links are basically treated as undirected. A visualization method considering the direction of the link has also been proposed. [5] In this study, we propose a visualization method considering the information of the hierarchy and circular structure of the network, and report the results applied to actual Japanese production network. The proposed method first decomposes the network into a potential flow network and a loop flow network by Helmholtz-Hodge decomposition. This decomposition determines the Helmholtz-Hodge potential corresponding to the hierarchical position in the network for each node of the network. Layout of visualization is determined by the spring electric model with adding constraints of coordinate corresponding to this Helmholtz-Hodge potentials. In the previous research [6], determining the layout corresponds to extracting the hierarchical structure, but in the proposed method, first of all, a hierarchical structure is uniquely obtained by Helmholtz-Hodge decomposition and then a layout according to it is determined.

Very recently, we have studied [3] structure of a Japanese production network with one million firms and five million supplier-customer links. We first constructed a directed network from the actual data of interfirm transaction relations and found that they form a tightly-knit structure with a giant strongly connected component surrounded by two half-shells constituting incoming-flow and outgoing-flow components for the core. The objective of this study is to advance the previous empirical analysis [3] on the industrial flow structure embedded in microscopic supplier-buyer relations with a special emphasis on its hierarchy and circularity. Hierarchy of the production network is expected to emerge from self-organization of supply chain in the industrial system. We also note that inner loops of production, giving rise to a nonlinear feedback mechanism to complicate dynamics of the industrial system, can be an engine for economic growth.

2 Methods

2.1 Helmholtz-Hodge decomposition

To delve further into the flow structure of a directed network, we take advantage of a mathematical tool called the Helmholtz-Hodge decomposition [2, 7]. It allows us to decompose flow on a directed network into a potential flow component and a circular flow component. In general, one can write flow F_{ij} running from node i to node j as

$$F_{ij} = F_{ij}^{(p)} + F_{ij}^{(c)} \quad (1)$$

The first term $F_{ij}^{(p)}$ on the right hand side of Eq. (1) denotes the potential flow from node i to node j which is given by

$$F_{ij}^{(p)} = w_{ij}(\phi_i - \phi_j) , \quad (2)$$

where ϕ_i is the Helmholtz-Hodge potential associated with node i and w_{ij} is a positive weight for linkage between nodes i and j . In the potential flow network, nodes are perfectly ranked; the potential flow thereby runs from a node with higher potential to a node with lower potential. On the other hand, the second term $F_{ij}^{(c)}$ denotes the circular flow component in which incoming flow and outgoing flow are exactly balanced at each node:

$$\sum_j F_{ij}^{(c)} = 0 , \quad (3)$$

so that there is no hierarchy among nodes in the circular flow network.

In practice, one can determine the potential ϕ_i for every node by minimizing the squared difference between the actual flow and the potential flow:

$$I = \frac{1}{2} \sum'_{i < j} w_{ij}^{-1} (F_{ij} - F_{ij}^{(p)})^2 \quad (4)$$

where the double summation excludes pairs of nodes which are not connected. Subtracting the potential flow thus determined from the actual flow leaves the loop flow. In addition, to remove arbitrariness in the potential determination, we impose the following condition on ϕ_i :

$$\sum_i \phi_i = 0 . \quad (5)$$

Here we assume that the flow structure of a directed network is given by

$$|F_{ij}| = \begin{cases} 1 & \text{(singly connected in one way)} \\ 0 & \text{(doubly connected in both ways)} \\ 0 & \text{(not connected)} \end{cases} \quad (6)$$

If volume of each transaction was available, we could this simplified flow structure with the actual one. Also we assume that the weight w_{ij} takes the following values depending on how the two nodes are connected:

$$w_{ij} = \begin{cases} 1 & \text{(singly connected in one way)} \\ 2 & \text{(doubly connected in both ways)} \\ 0 & \text{(not connected)} \end{cases} \quad (7)$$

The Helmholtz-Hodge potential of nodes in a directed network identifies their hierarchical positions in the flow structure. In contrast, the circular flow component illuminates feedback loops built in the system.

2.2 Flow based community detection

Community detection is widely used to elucidate structural properties of large-scale networks. In general, real networks are highly non-uniform. Community detection singles out groups of nodes densely connected to each other in a network to divide it into modules. This enables us to have a coarse-grained view on structure of such complicated networks. The map equation method [9] is one way to detect communities in a network. This method is found to be one of the best performing community detection technique when compared with others [8]. It is a flow-based and information-theoretic method depending on the map equation defined as

$$L(C) = q_{\curvearrowright} H(C) + \sum_{i=1}^m p_{\circlearrowleft}^i H(\mathcal{P}^i). \quad (8)$$

Here $L(C)$ measures the per step average description length of dynamics of a random walker migrating through links between nodes of a network with a given node partition $C = \{C_1, \dots, C_\ell\}$ and consists of two parts. The first term arises from movements of the random walker across communities, where q_{\curvearrowright} is the probability that the random walker switches communities and $H(C)$ is the average description length of the community index codewords given by the Shannon entropy. The second term arises from movements of the random walker within communities, where p_{\circlearrowleft}^i is the fraction of the movements within community C_i and $H(\mathcal{P}^i)$ is the entropy of codewords in module codebook i .

If the network has densely connected parts in which a random walker stays long time, one can compress the description length of the random walk dynamics on a network by using a two-level codebook for nodes adapted to such a community structure, an analogy to geographical maps in which different cities recycle the same street names such as main street. Therefore, obtaining the best community decomposition in the map equation framework amounts to searching for the node partition that minimizes the average description length $L(C)$. The code of the map equation algorithm is available at <http://www.mapequation.org>.

2.3 Visualization based on a spring-electric model

A spring-electric model is adopted in which pairs of nodes with direct relations are physically connected by springs, and nodes in any pair repel each other through a repulsive Coulomb force. The attractive force of the spring keeps intimate nodes close in space. On the other hand, the repulsive Coulomb force tends to distribute firms uniformly over the available space and prevents entanglement of a network. We then take full advantage of a molecular dynamics (MD) method [1, 4] for an optimized configuration of nodes in the model. The ground state in the model is a leading candidate for this configuration. The MD simulation works well to reproduce an ordered structure, with the lowest-energy forms such as crystals of materials generated through slow cooling, starting from any initial configuration. We expect that the simulation is also successful in visualizing the network.

The interaction force between nodes l and m for the model is explicitly written as

$$F(r_{lm}) = -k_{lm}r_{lm} + \frac{q_l q_m}{r_{lm}^2}, \quad (9)$$

where k_{lm} is the spring constant for the attraction between the nodes. If the nodes are directly connected, $k_{lm} = k$; otherwise $k_{lm} = 0$. q_l denotes the Coulomb charge for node l . Here we neglect the direction of links (flow of goods or money) and assume that q_l and k_{lm} take on identical values for every node and pair, respectively.

3 Results and Discussion

The present analysis is based on a big data of 4,974,802 transaction relations between 1,066,037 firms in Japan which was collected by the Tokyo Shoko Research, Ltd. in 2016.³ This data virtually covers whole industrial activities in Japan. We regard firms as nodes and transaction relations between them as directed links spanning from suppliers to customers to construct the latest production network in Japan. Since information on the volume of each transaction is not available, we assume that all the links have the same weight.

To elucidate flow structure in the TSR transaction network, we begin with a bow-tie decomposition of the network as has been widely used to understand the structure of various complex networks including the world wide web and metabolic networks. The decomposition classifies nodes in a directed network according to the way in which they are mutually connected: IN component, GSCC (Giant Strongly Connected Component), OUT component, and others. The GSCC is the largest group of nodes in which any pairs of nodes are connected bidirectionally. The IN component is a collection of nodes which have a path to the GSCC, but no reverse path to come back. The OUT component is defined in the other way around, that is, a collection of nodes which are reachable only from the GSCC. The TSR transaction network is decomposed into 219,927 IN components, 530,174 GSCC components, 278,880 OUT components and 37,056 Others.

We obtained an optimized layout of the network in three-dimensional space with information of the Helmholtz-Hodge potential obtained for individual nodes. The results visualized in Fig 1 and 2. Nodes are aligned in the z direction according to their values of the Helmholtz-Hodge potential; basically, transaction flows from top to bottom. On the other hand, the x and y coordinates of nodes are determined by minimizing the potential energy in a spring-electric model.

The hierarchical flow is dominant in the IN component, which has mainly one-way flow to the GSCC out of its definition. This is also true for the OUT component. On the other hand, the GSCC has more complicated flow structure; both hierarchical and circular flow components coexist in it. For the purpose of

³ This is the largest connected component in the network obtained from the original data, containing 99.3% of all active firms listed in the data.

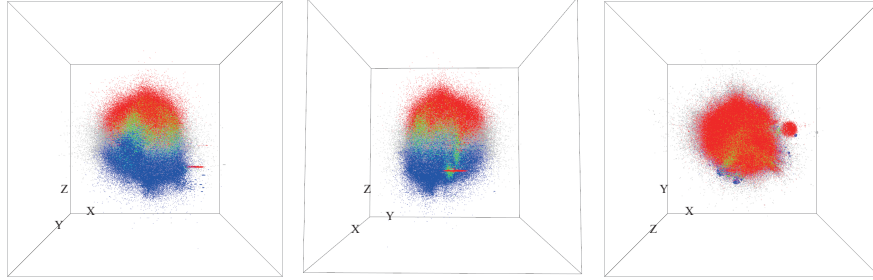


Fig. 1. The IN component (red), GSCC (green), and OUT component (blue) of the TSR transaction network visualized in three-dimensional space. Nodes are aligned in the z direction according to their values of the Helmholtz-Hodge potential; basically, transaction flows from top to bottom. On the other hand, the x and y coordinates of nodes are determined by the energy minimum principle with a spring-electric model.

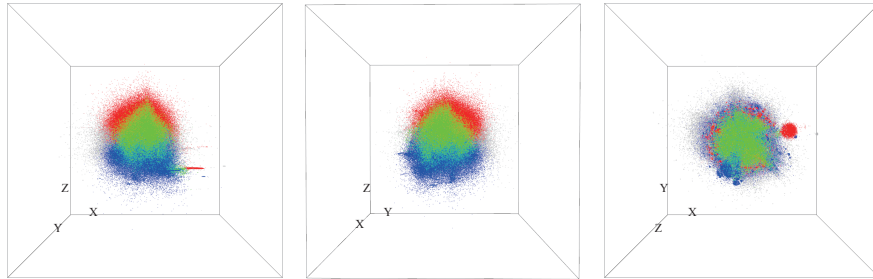


Fig. 2. Half-cut cross sections of the 3D views of the TSR networks as shown in Fig. 1.

this study, therefore, we concentrate on the flow structure of the GSCC, especially its circularity. To identify important loop structure in the TSR transaction network, we apply the map equation method for community detection to the circular flow network which is obtained from GSCC of the TSR transaction network by Helmholtz-Hodge decomposition. The total number of communities is 18,660 and the largest community has approximately 5,000 firms. These communities are dense parts of the circular structure in the network. The 10 largest communities are illuminated in Fig. 3 with the same node configuration as in Figs. 1 and 2. Nodes are aligned in the z direction according to their values of the Helmholtz-Hodge potential.

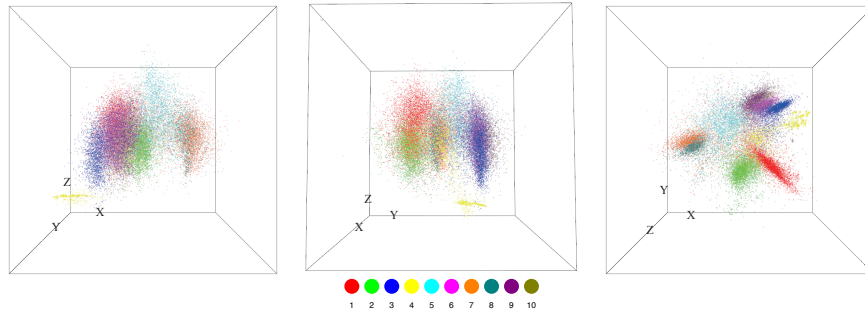


Fig. 3. The 10 largest communities in the circular flow network on the GSCC of the transaction network, visualized in three-dimensional space with three different points of view. The same configuration of firms is used as in Fig. 1.

One can characterize these 10 communities by industrial and regional affiliations of their constituent firms. They are divided into two contrastive groups. The first, second, fourth and fifth largest communities are featured by manufacturing and wholesales industries; medical, health care & welfare industry is additionally important for the fourth community. On the other hand, the remaining 6 communities are featured by construction industry. Also all the major communities have prominent regional characteristics. The manufacturing and wholesales dominant communities are basically metropolitan communities except for the second largest community in which Hokkaido and some provincial prefectures play a key role. In contrast, distribution of the regional affiliations in the construction dominant communities are well localized at prefecture level.

The community 1, 2, 4, and 5 are all communities which manufacture and wholesales dominates, but one can extract difference among them by using more detailed classification. The community 1 includes many manufacture and wholesales of textile and apparel. The community 2 includes fisheries cooperative, wholesales and retail trade of seafood, manufacture of food of seafood. The community 4 includes Medical and health services, manufacture and wholesales of

pharmaceutical products. Most of medical and health services are general hospitals and clinics. The community 5 includes many manufacture and wholesales of metal products and construction. In this way one can characterizes manufacture and wholesales communities by product which well related community. Although the community is an extracted dense part of loop flow network, it includes manufacture, wholesales and retail trade and form hierarchical structure of product generally called supply chain. The community 3, 6, 7, 8, and 9 are communities which construction dominates. Although industry distribution of the construction communities similar to each other, it can be characterized by its locality. In addition, firms contributing to the flow from downstream to upstream are not the main industries of the supply chain, but are complementary industries (road freight transport, equipment installation work, etc.).

From these results, we found that within the community there are a hierarchical supply chain consisting of main industry of the community and a circular structure mainly consisting of industries other than the main industry.

4 Conclusion

The comprehensive dataset of interfirm transaction relations in Japan enabled us to study the industrial flow structure of the nations production network with a sound microscopic foundation. Especially we emphasized its hierarchy and circularity. By adopting the Helmholtz-Hodge decomposition, we separated the flow structure of the GSCC of the transaction network into two components: potential flow and circular flow. The potential flow between a pair of firms is given by difference of their potentials, and hence, the potential of a firm identifies its hierarchical position in the transaction network. On the other hand, the circular flow component illuminates feedback loops built in the network. The layout was calculated and visualized by the spring electric model with the constraint condition corresponding to the Helmholtz-Hodge potential. We also identified dominant clusters of firms forming feedback loops by applying the map equation method to the extracted circular flow network. We found that both hierarchical and loop structure coexist within the major industries such as construction, manufacturing, and wholesales.

References

1. Allen, M., Tildesley, D.: Computer simulation of liquids (1987)
2. Bhatia, H., Norgard, G., Pascucci, V., Bremer, P.T.: The Helmholtz-Hodge decompositiona survey. *IEEE Transactions on visualization and computer graphics* **19**(8), 1386–1404 (2013)
3. Chakraborty, A., Kichikawa, Y., Iino, T., Iyetomi, H., Inoue, H., Fujiwara, Y., Aoyama, H.: Hierarchical communities in walnut structure of japanese production network (2018), available at SSRN: <https://ssrn.com/abstract=3129974> or <http://dx.doi.org/10.2139/ssrn.3129974>
4. Frenkel, D., Smit, B.: Understanding molecular simulation: From algorithms to applications (2002)

5. Fujita, Y., Fujiwara, Y., Souma, W.: Large directed-graph layout and its application to a million-firms economic network. *Evolutionary and Institutional Economics Review* **13**(2), 397–408 (2016)
6. Hu, Y.: Efficient, high-quality force-directed graph drawing. *Mathematica Journal* **10**(1), 37–71 (2005)
7. Jiang, X., Lim, L.H., Yao, Y., Ye, Y.: Statistical ranking and combinatorial Hodge theory. *Mathematical Programming* **127**(1), 203–244 (2011)
8. Lancichinetti, A., Fortunato, S.: Community detection algorithms: a comparative analysis. *Physical review E* **80**(5), 056117 (2009)
9. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* **105**(4), 1118–1123 (2008)