Detangling the multilayer structure from an aggregated network

Aobo Zhang, An Zeng*, Ying Fan and Zengru Di
School of Systems Science, Beijing Normal University, Beijing, 100875, People’s Republic of China
* Author to whom any correspondence should be addressed.
E-mail: anzeng@bnu.edu.cn

Keywords: multilayer networks, network splitting, international trading networks, link clustering coefficient

Abstract
Multiplex interactions are common and essential in real-world systems. In many cases, we can only obtain aggregated networks without detailed information regarding the type of links contained within. Such single-layer networks oversimplify the structural information and lead to misunderstandings of some properties of real systems. In this context, network splitting which aims to correctly separate an aggregated network into multilayer networks, is a meaningful problem to address. To this end, we propose a simulated-annealing-like algorithm based on the link clustering coefficient. We verify the validity of this algorithm with several synthetic networks. Inter-similarities of layers are also taken into consideration, and we can find that the proposed method is valid even if there is a certain proportion of overlapping links between layers. Finally, we apply the algorithm to real international trading networks, which results in accurate splits of different layers.

1. Introduction
Networks can be used to characterize the structure of many complex systems and provide powerful representations of the interactions between agents. The relations among social individuals [1], biological interactions [2], traffic routes, and many other circumstances [3, 4] can be represented as networks which can reveal the properties from the structural information. Such networks consist of only a single type of entity. However, the networks that neglect the types of relations oversimplify the complexity of the systems’ structural properties. In most social and natural systems, there is a set of entities among individuals who can represent the multiple types of interactions [5–7]. Therefore, it is important to use multilayer networks to construct such complex systems with multiple types of connections or subsystems. In social networks, the relationships between people can be composed of many interactive relationships, such as families, friends, and co-workers [1, 8]. A brain network that consists of structural and functional layers is potentially more informative than either single-layer network on its own [9, 10]. The results of previous studies have demonstrated the differences in robustness between networks of networks and aggregated networks under attack [11, 12]. Ignoring the multiplex of networks results in information loss and leads to inaccurate interpretations.

Currently, the study of the complexity of multilayer networks is an essential research direction, and intense efforts have been made to investigate multiplex systems [13]. Multilayer networks can disclose more attributes of complex systems than single-layer networks [14, 15]. Research has demonstrated the various results obtained when analyzing a system from a multilayered perspective [16]. The results showed that neglecting multiple layer can result in incorrect identification [17]. There are many research works that has focused on multilayer networks. For example, frameworks for calculating node centrality have been proposed to find the central roles when considering different types of relations in different layers [18]. Various types of dynamic processes in multilayer networks have also been discussed, and one such study provided a useful method for improving the understanding of the complexity in a system. In addition, the existence of multiplex layers leads to a drastic impact on the dynamics of all overall systems [19–22]. Some
of the interactions in of multilayer network representations are considered to be redundant and uninformative [23]. In this situation, some layers can be aggregated [24] while maximizing the distinguishability of the remaining layers based on quantum theory and retaining as much information as possible [17]. Another framework for reduction involves aggregates the system layers by using eigenvector-based centrality [25]. For link prediction in multilayer networks, the helpful information contained in the link from one layer to another gives us a principled way to define interdependence between layers and compress redundant information [26].

The studies referenced above assumed that the interactions in each layer were known, but we may not have complete information of the relations in many real systems. Therefore, there is still an important problem that has not been studied well. Sometimes, we can only obtain an aggregated network without knowing the types of relationships within. Such networks that neglect the existence of different interactions between individuals may lose much information. Taking the social networks as an example, the relationship between people in the social network may include friend relationship, family relationship, etc. These relationships are naturally modeled by a multi-layer network where each type of relationship consists of a network layer. Therefore, it is essential to find a method to distinguish between different types of relations. In this paper, we propose a framework for splitting an aggregated network into a multilayer network.

In this paper, we aim to address the question of how to develop a universal framework for splitting aggregated networks with no extra supporting structural information. To solve this problem, we propose a simulated-annealing-like algorithm based on the link clustering coefficient to split one aggregated network into a two-layer network with different types of relations in each layer separately. The method is validated with both artificial networks and real networks (e.g. international trading networks). The results show that our method can perform well for networks with high clustering coefficients. We also consider the overlap ratio of the two layers and demonstrate the validity of our method for networks that do not highly overlap. Experimental results with respect to two applications on real systems corroborate the validity of the proposed approach.

2. Method

In this paper, our aim as shown in figure 1 is to split an observed network \( C \) into two networks; \( A \) and \( B \) represent the separated layers with the same type of relationship in each network. To solve this problem, we define an objective function (see equation (1)) which is the minimum link clustering coefficient between the networks \( A \) and \( B \). Note that the link clustering coefficient of link \( ij \) is defined as \( \frac{N_{ij}}{k_i \ast k_j} \) where \( N_{ij} \) is the number of common neighbors between nodes \( i \) and \( j \), and \( k_i \) and \( k_j \) are the degrees of node \( i \) and \( j \), respectively. The basic idea is that if edges are split correctly, the link clustering coefficient in each layer must be higher than that in the case of wrong split. The link clustering coefficient can be regarded as a measure of the similarity between nodes. If the similarity between two nodes is higher, there will be a larger probability of these two nodes being connected in this layer. We adopt this definition method because it can judge whether the edge should exist in this layer by measuring the similarity of the nodes. Such an objective function can avoid the situation where one network is very dense while the other is almost empty. The goal is to maximize the objective function.

\[
F = \min (\text{Cluster Coeff}(A), \text{Cluster Coeff}(B)).
\] (1)

Then we use a simulated-annealing-like algorithm to optimize the objective function. Initially, \( C \) is randomly split into two networks \( A \) and \( B \) with an equal number of links. During each step, we randomly select one network (\( A \) or \( B \)) and choose one link; this link is then placed in the other network. For example,
Figure 2. The objective function and number of errors versus the number of steps. Both networks A and B are regular networks with $N = 100$. (a) and (b) show the results of network with $k_A = 4$ and $k_B = 4$. (c) $k_A = 8$ and $k_B = 8$. (d) $k_A = 4$ and $k_B = 8$.

Figure 3. The evolution of the errors and objective function during the optimization process when separating two SW networks. The network parameters of A and B are both $N = 100$ and $k = 4$. The mixed network C is obtained by $C = A + B$. (a)–(c) The results of the random split with initial configurations of $p = 0, 0.1$ and 0.3. (d)–(f) The results of the true split with initial configurations of $p = 0, 0.1$ and 0.3.

if A is selected, we choose a link from A randomly and move it to network B. This modification is accepted under two conditions: (1) the objective function is improved; (2) a randomly generated number (between [0, 1]) is smaller than $q$ which is a value that decreases as steps increases, where $q = 1$ at first and $q = q \times 0.999$ after each step. At the beginning, $q$ is large, so the changes in links are quite random. As the number of iterations $q$ decreases, the modifications depend more on condition (1) than on (2). This technique can help the optimization process jump out of local minima. In order to avoid a situation where the degree of many nodes in one layer of network is zero, we introduce a penalty term $(\sqrt{(1 - n_0)}/N)$, where $n_0$ is the number of nodes whose degree is zero) to punish the objective function in such situations. Eventually, we hope to obtain networks A and B that contain the same type of interaction. The proposed method splits the aggregated network based on the clustering of links and the assumption of high levels of clustering in real networks.

3. Results

3.1. Simulation in synthetic networks

To begin our analysis, we consider regular networks (nearest-neighbor networks) in which each node connects to the $k$ nodes with the closest IDs to its own. To obtain an aggregated network with two types of
relations, we generate two networks (with $N = 100$ and $k = 4$), and then combine them into one network as the observed network $C$. Note that, when we combine $A$ and $B$ to obtain the mixed network $C$, the ID of nodes in both $A$ and $B$ are randomly reordered. This process can be regarded as adding the links in network $B$ to network $A$. As the ID of nodes in these two networks are randomly reordered, a short range link in $B$ may connect two very distant nodes in $A$ when adding links of $B$ to $A$. To simplify the problem, we remove all overlapping links. The results are shown in figures 2(a) and (b). We calculate the number of error links to demonstrate the validity of this algorithm. We can see that the error decreases as the number of steps increases and finally almost reaches zero. We then consider a denser network (with $N = 100$ and $k = 8$) and a more complicated case where $A$ and $B$ have different densities (with $k_A = 4$ and $k_B = 8$). The results are still robust under such situations.

Next we try to apply this method to small-world (SW) networks with different probabilities $p$ [27]. Network layer $A$ is an SW network with $N = 100$, $\langle k \rangle = 4$, and a tunable $p$ (rewiring probability). Network layer $B$ is also an SW network with $N = 100$, $\langle k \rangle = 4$, and a tunable $p$ (rewiring probability). Then we obtain the observed mixed network $C = A + B$. Note that, when we combine $A$ and $B$ to obtain $C$, the IDs of the nodes in both $A$ and $B$ are randomly reordered. The results are shown in figure 3. The error decreases as the number of steps increases when $p$ is small. If $p > 0.1$ (see figure 3(c)), the accuracy starts to decrease, because the clustering property is destroyed when $p$ is too large (see more details in section S1 (https://stacks.iop.org/NJP/23/073046/mmedia) in the supplementary materials). We also try to use the true
split as the initial configuration during the first step and the error does not have fluctuate greatly with the number of steps (see figures 3(d) and (e)). This is because if the objective function is effective, the true split of the aggregated network should correspond to the optimal objective function. This also reveals that it is suitable to use the link clustering coefficient as the objective function.

To further support our findings, we apply our method to different SW networks with different values of $p$ in figure 4. The error increases with $p$ and $p = 0.1$ is a critical value. The method performs well when $p < 0.1$. A high clustering coefficient is the basic characteristic of SW networks, so we believe that it is an important factor that influences the validity of the proposed method. The results verify this relationship (see figures 4(c) and (d)). Three of the real cases are also consistent with the conclusion in section S2 of supplementary materials. (We disorder the nodes to obtain a new network $B$ and consider the original network as $A$.) According to figure 4, the trend of the two initial configurations (random split and true split) are similar. We want to know if we can use the true split to prove the validity of the simulated-annealing-like algorithm. Then we contrast the results of the two initial configurations in figure 4 and there is no obvious difference.

We extend this method to a three-layer network, and calculate the number of wrong edges and the values of objective function. The three-layer network consists of three regular networks. The aggregated network is obtained by combining these three networks by randomly reordering the ID of the nodes in each layer. We find that the three layers can be well identified by our method. The results are shown in figure 5.

### 3.2. Layer inter-similarity

In reality, one node may interact with others in layers $A$ and $B$ simultaneously. In this section, we try to perform the simulation under different layer inter-similarities between the two layers. To solve this problem, we construct networks with different layer inter-similarities. Considering a network with $N$ nodes, we choose a certain ratio of nodes to maintain the interactions and structure among these nodes and define the ratio $Q$ as the index for measuring the layer inter-similarity. Then, we disorder the nodes that are not chosen, so the relations among them are quite different than those in the original network. We cannot deal with overlapped links in our proposed method, but there are more overlapped links as $Q$ increases.
Therefore, we place them into the two networks randomly. Compared with the method that removes all the overlapped links, the random method has less influence on the densities of the networks if the ratio $Q$ is high. After the above process is completed, we can obtain two networks $A$ and $B$ and aggregate them into one network $C$. The application results of the algorithm for such networks with different values of $Q$ are shown below. Figure 6(a) is an example of regular networks ($N = 100$ and $k = 4$) with $Q = 0.1$, and we can see that the error decreases to a low value as the number of steps increases. To explore this phenomenon further, we try to determine the influence of $Q$ in figure 6(b). Larger values of $Q$ result in higher numbers of errors. We also study the validity of the algorithm in SW networks in figures 6(c) and (d) (random split and true split as the initial configurations, respectively). The results show that our algorithm can separate the network even with a relatively high layer inter-similarity. We also consider other methods to deal the overlapping links in section S3 in the supplementary materials.

### 3.3. Application in real networks

To verify the algorithm further, we apply it to real-world systems in figure 7 (international trading networks) [28]. In the results of the international trading networks, we can separate almost 70% of the lines, and this is much better than the results of random split. The two layers are the export relations of different kinds of products, including ‘vegetables & fruits’ and ‘cultural & artware products’ such as books and art. We calculate the revealed comparative advantage (RCA) index values [29] of all countries and products according to the export volumes and retain the export information when the RCA is above 1. The layer of one product is established according to the remaining export relationship. We try to distinguish the relationships of different products. In figure 7(a), we show the performance of the algorithm when we split the aggregated network (international trading network) into two networks, with each color representing one type of export product. Network $A$ is the layer of ‘vegetables & fruits’, and network $B$ represents the layer of ‘cultural & artware products’. In layer $A$, most of the links are brown and the green lines arranged incorrectly are relations belonging to layer $B$. A similar situation can be seen in layer $B$. In figures 7(b) and (c), we show the proportion of error lines versus the variation in the objective function with the random and true configuration respectively. In addition, we try to separate other aggregated networks for different kinds of products. In table 1, we show the detailed values of proportion of errors lines, final objective function and objective function of true split. We also show the other two examples in section S4 in supplementary materials. Although the objective function of the true split does not correspond to the
Table 1. We apply the algorithm to different networks for other products such as ‘fossil fuels’, ‘fabrics materials’. The splitting results in terms of the proportions of error lines and final objective functions are shown in the table. We also compare these with the objective function of the true split.

| Aggregation Network                                      | Proportion of error lines (%) | Final objective function | Objective function of true split |
|----------------------------------------------------------|-------------------------------|--------------------------|----------------------------------|
| Vegetables and fruits-metals                             | 35.31                         | 0.3109                   | 0.4771                           |
| Vegetables and fruits-fabrics materials                  | 38.01                         | 0.4308                   | 0.3987                           |
| Vegetables and fruits-cultural and artware products       | 35.70                         | 0.3714                   | 0.3987                           |
| Wine and tobacco-clothing articles                       | 12.84                         | 0.5003                   | 0.4867                           |
| Wine and tobacco-fossil fuels                            | 21.64                         | 0.5262                   | 0.5021                           |
| Wine and tobacco-fabrics materials                       | 38.83                         | 0.4635                   | 0.4086                           |
| Clothing articles-fossil fuels                           | 29.83                         | 0.4968                   | 0.4296                           |
| Metals-iron and steel products                           | 39.91                         | 0.4632                   | 0.3974                           |
| Fabrics materials-cultural and artware products          | 29.66                         | 0.4455                   | 0.3006                           |

Optimum according to the results, the proportion of error lines can be reduced to nearly 30\% relative to the random split. This means that we can separate almost 70\% of the relations between each pair of products by the proposed algorithm. We also apply our method in a social network consisting of two layers. The layers are respectively friend relationship in Facebook and work relationship [30] obtained from questionnaires. The two layers of networks are merged into an aggregated network. One can also see that the proportion of error links can be reduced to nearly 36\% (see figure S9 in SM).

4. Discussion

The separation of an aggregated network is a very important problem from both theoretical and practical points of view. Currently, larger and more detailed data-sets than ever before are being generated at an increasingly fast rate, and they are considered to be highly essential in many aspects of society. The increasing supply of information provides many materials for analysis, and we can therefore obtain a deep understanding of the structures of different complex systems and mechanisms. There are studies that discuss the reducibility of multilayer networks to reduce the amount of unnecessary information. The problem we are concerned with in this paper is that sometimes we can obtain only one aggregated network but not the detailed relationships among the agents contained with. For example, in some social application networks, the information we obtain is only the relationships of ‘following’ and ‘followed’, but the real interaction (friends, family or co-workers) among them are not shown. We believe that it is useful to separate networks for targeted recommendation and spreading control. Hence, the idea of providing a method to split such aggregated networks is proposed.

In this paper, we focus on such problems and propose a simulated-annealing-like algorithm. The basic idea is to maximize the link clustering coefficient of the two separated networks by changing links. We apply the proposed method for several synthetic networks, such as regular and SW networks. The results demonstrate the feasibility of the algorithm. We find that the validity is influenced by the clustering coefficient between the networks and that the algorithm can only perform well in networks with high coefficients. The layer inter-similarity $Q$ is also considered. We can use the method to solve problems as long as $Q$ is not too high. The results in the real-world systems also show the credibility of the proposed algorithm.

We consider multilayer networks with two layers in this paper. The proposed method cannot perform well in all networks. Networks with higher clustering coefficients lead to better performances. Moreover, there are various structural properties that affect the performance of our method. Since our method relies on only the structural information to split a network into different layers, it has several limitations, especially when facing some specific structures. Firstly, the method cannot work in a very sparse network, as such network cannot provide enough structural information (link clustering) for the method to decide which layer a link should belong to. Secondly, the method requires a relatively low overlap of the links in different layers, as we shown in figure 6. Finally, the degree heterogeneity may decrease the accuracy of the method. In this case, some links of the hubs may be placed in the wrong layers. Although our approach is not exempt from limitations, it highlights the importance of detangling the multilayer structure from the observed aggregated networks. It opens a new research direction for analyzing and understanding real networks. We hope to try to improve its applicability in the future and apply this method to increasingly complicated systems.
Acknowledgments

The authors would like to thank Dr. Matus Medo for the helpful discussion. This work is supported by the National Natural Science Foundation of China (NSFC) (Grant No. 71731002).

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://comtrade.un.org/.

References

[1] Wasserman S and Faust K 1994 Social Network Analysis (Methods and Applications) vol 8 (Cambridge: Cambridge University Press)
[2] Shen-Orr S S, Milo R, Mangan S and Alon U 2002 Nat. Genet. 31 64–8
[3] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D U 2006 Phys. Rev. 424 175–308
[4] Newman M 2010 Networks (An Introduction) (Oxford: Oxford University Press) https://books.google.com/books?id=LrFaU4XCxOcC
[5] Kivelä M, Arenas A, Barthelemy M, Gleeson J P, Moreno Y and Porter M A 2014 J. Complex Netw. 2 203–71
[6] Boccaletti S, Bianconi G, Criado R, Del Genio C I, Gómez-Gardeñes J, Romance M, Sendiña-Nadal I, Wang Z and Zanin M 2014 Phys. Rep. 544 1–122
[7] Moreno Y and Perc M 2019 New J. Phys. 22 010201
[8] Scott J 2012 Social Network Analysis (London: SAGE Publications) https://books.google.com/books?id=MJcdGBYDGE
[9] De Domenico M 2017 Giga Science 6 1–8
[10] Sun Y and Han J 2013 ACM SIGKDD Explor. Newsl. 14 20–8
[11] Gao J, Buldyrev S V, Havlin S and Stanley H E 2011 Phys. Rev. Lett. 107 195701
[12] Dong G, Gao J, Du R, Tian L, Stanley H E and Havlin S 2013 Phys. Rev. E 87 052804
[13] D’Agostino G and Scala A 2014 Networks of Networks: The Last Frontier of Complexity vol 340 (Berlin: Springer)
[14] Jeub L G, Mahoney M W, Mucha P J and Porter M A 2015 arXiv:1510.05185
[15] Faqeeh A, Osat S and Radicchi F 2018 Phys. Rev. Lett. 121 098301
[16] Szell M, Lambiotte R and Thurner S 2010 Proc. Natl Acad. Sci. 107 13636–41
[17] De Domenico M, Nicosia V, Arenas A and Latora V 2015 Nat. Commun. 6 6864
[18] Solá L, Romance M, Criado R, Flores J, García del Amo A and Boccaletti S 2013 Eigenvector centrality of nodes in multiplex networks Chaos 23 033131
[19] Zuzek L G A, Stanley H E and Braunstein L A 2015 Sci. Rep. 5 12151
[20] Pan L, Wang W, Cai S and Zhou T 2019 Phys. Rev. E 100 022316
[21] Danziger M M, Bonamassa I, Boccaletti S and Havlin S 2019 Nat. Phys. 15 178–85
[22] Chen D, Shi D D, Qin M, Xu S M and Pan G J 2018 Phys. Rev. E 98 012319
[23] Taylor D, Shai S, Stanley N and Mucha P J 2016 Phys. Rev. Lett. 116 228301
[24] Petrov T and Tognazzini S 2020 Centrality-preserving exact reductions of multi-layer networks Int. Symp. Leveraging Applications of Formal Methods (Springer) pp 397–415
[25] De Bacco C, Power E A, Larremore D B and Moore C 2017 Phys. Rev. E 95 042317
[26] Watts D J and Strogatz S H 1998 Nature 393 440–2
[27] https://comtrade.un.org/
[28] Balassa B 1965 Manch. Sch. 33 99–123
[29] Magnani M, Micenková B and Rossi L 2013 Combinatorial analysis of multiple networks (arXiv:1303.4986)