Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P013-cd

Community Detection Algorithm for Natural Gas Pipeline Network Based on Transmission Characteristics

Heng Ye

School of Energy Resources, China University of Geosciences (Beijing), China. PipeChina Oil & Gas Control Center, Beijing, China. E-mail:heisyh@163.com

Zhiping Li

School of Energy Resources, China University of Geosciences (Beijing), China. E-mail: lzpoffice@126.com

Yu Li

PipeChina Oil & Gas Control Center, PipeChina, China. E-mail: 825481000@qq.com

Yiran Liu

PipeChina Oil & Gas Control Center, PipeChina, China. E-mail: yiran960406@hotmail.com

Jiale Liu

Department of Marketing, PipeChina, China. E-mail: liujl@pipechina.com.cn

Enrico Zio

Department of Energy, Polytechnic University of Milan, Italy. E-mail: enrico.zio@polimi.it

Huai Su

School of Storage and Transportation Engineering, China University of Petroleum (Beijing), China. E-mail: suhuai@cup.edu.cn

The community detection is beneficial for natural gas pipeline network to optimize daily operation and divide regions in a reasonable way. However, traditional community detection methods only focus on the network topology structure and cannot reflect the transmission characteristics of natural gas pipeline network. To solve this problem, a gas flow tracing algorithm, based on the proportional sharing principle, is applied to determine the contribution of each source node to each download node. According to the gas flow tracing result, the gas transmission correlation strength is defined, the gas transmission modularity is constructed, and their physical meaning is elaborated. By replacing the traditional modularity with the gas transmission modularity, we improve the fast greedy community algorithm to form a new community detection algorithm for natural gas pipeline network, which can automatically obtain the optimal community division under different working conditions. The case study in this paper demonstrates how our method adaptively identify and divide communities for different conditions based on gas transmission characteristics.

Keywords: Natural gas pipeline network, community detection, gas flow tracing, gas transmission modularity.

1. Introduction

Natural gas pipeline network is a complex transmission network, which is usually composed of multiple gas sources, download or distribution stations, interconnection stations and compressor stations. Similar to the power grid and railway network, the natural gas pipeline network also has the characteristics of high degree of interconnection, uneven distribution of flow, and locally dense or sparse structure. It shows a strong "community structure" feature, which means the internal connection within community is close, while the connection between communities is relatively sparse.

In 2002, Girvan and Newman proposed the concept of community structure for social and biological networks, and then established the GN algorithm for community detection[1]. Subsequently, different community detection algorithms developed and gradually evolved into fast community detection algorithm based on the maximum modularity[2], multi level recursive bisection algorithm[3], muti agent algorithm[4], K-means cluster algorithm[5], etc. Based on the preset number of communities, the GN algorithm deletes the edges in order of the edge betweenness from large to small until the original network is divided into preset number of sub-networks[6]. GN algorithm is simple and direct, but it is time-consuming and not suitable for large networks because the betweenness of all edges needs to be calculated in each step. Newman optimized the GN algorithm and proposed a community detection algorithm based on the maximum modularity[7]. He used modularity as the evaluation index to automatically determine the number of divided communities through continuous aggregation optimization. With the improvement of data structure and calculation process, Newman then proposed the fast greedy community algorithm (FGC algorithm), which further increases the calculation speed and is more suitable for large-scale network analysis[8].

The above algorithms have good versatility and can be quickly applied to various networks with good result. However, these algorithms are based on pure network topology structure, lacking the consideration of network function. For the natural gas pipeline network, this paper proposes a set of community detection algorithm that can reflect the gas transmission characteristics. It may provide scientific basis for the division of dispatch area of control center, the location of regional maintenance center or emergency repair center, and other similar decision related to daily production operation.

2. An algorithm for gas flow tracing

It is obviously that the main purpose of natural gas pipeline network is to transport natural gas. From the perspective of complete hydraulics, the correlation between nodes mainly depends on the supply relationship between source nodes and download nodes. In fact, the natural gas pipeline network is highly interconnected and sometimes it is difficult to distinguish the source of natural gas downloaded at download node. Taking the West-east Gas Pipeline in China as an example, along the main trunk, the gas from starting node will be mixed 14 times by other gas source from interconnected pipelines before reaching the end node. Therefore, before conducting community detection based on gas transmission characteristics, an algorithm for gas flow tracing is needed to determine the correlation between nodes.

Currently, there is no clear algorithm for gas flow tracing. Since the natural gas pipeline network is a typical complex network with the physical foundation for applying the complex network theory, similar power grid and corresponding algorithms can be used as reference. Generator node and load node in power grid can be seen as source node and download node in gas pipeline network.

The power grid was gradually separated from power plants or power suppliers and operated independently in the 1980s. Later, Bialek first proposed a power flow tracing algorithm according to the complex network theory and lossless flow treatment. The algorithm was applied to finds the power contribution relationship between every pair of generator node and edge, every pair of edge and load node, every pair of load node and generator node[9,10]. Operators of power grid can use this algorithm to determine the power distribution among transmission lines and the utilization degree of each transmission line[11]. Kirschen proposed another flow tracing algorithm to determine the contribution of every generator node to load node group[12].

The above two flow tracing algorithms of power grid are based on the same two core assumptions, which are also applicable to natural gas pipeline network. First, it is assumed that power will be preferentially supplied to the load node closest to the generator load. In natural gas pipeline network, it can also be assumed that natural gas preferentially supplies the download node closest to the source node. Secondly, the proportional sharing principle was constructed. At a certain node in power grid , the principle assumes the inflow power is fully mixed before being transmitted out by each outflow edge. It means each unit power leaving the node contains the same proportion of the inflow power as the total nodal power[13]. In natural gas pipeline network, natural gas from different gas sources is fully mixed at the pipeline interconnection node, which is consistent with the proportional sharing principle.

In summary, given the throughput of every edge and node, the flow tracing algorithm of Bialek can be used in natural gas pipeline network, which figures out how the total upload volume at each source node is assigned to the download volume at each download node.

3. An algorithm for community detection based on gas transmission characteristics

The fast greedy community algorithm (FGC algorithm) proposed by Newman uses the modularity to quantitatively evaluate the community detection[8]. The calculation and iteration process will be guided by the greedy algorithm until the maximum modularity is obtained[14]. The modularity can be calculated by the following formula.

$$Q=\frac{1}{2m}\sum_{ij}\left[A_{ij}-\frac{k_i k_j}{2m}\right]\delta(c_i,c_j) \qquad (1)$$

Here, A_{ij} is an element in the network adjacency matrix. The adjacency matrix is a square matrix of order n, and n is the total number of nodes in network. When node i and node j have an edge directly connected, A_{ij}=1, otherwise A_{ij}=0. Since node i has no edges directly connected with itself, A_{ii}=0. k_i and k_j are the degree of node i and node j respectively; m is the total number of edges in network and its expression is as follow.

$$m=\sum_{ij}A_{ij}/2 \qquad (2)$$

c_i and c_j represent the community where node i and node j are located. If node i and node j are in the same community, $\delta(c_i,c_j)=1$. Otherwise, $\delta(c_i,c_j)=0$.

As mentioned before, the FGC algorithm only considers the network topology structure and simply divides nodes that have close topological connection into one community. If it is applied directly to the gas pipeline network, the divided communities may not include a complete hydraulic system, which means some communities may consist of download nodes only. This result may break the complete transmission process from source nodes to download nodes. Besides, when the important gas transmission characteristic of source supply and user demand changes, the community detection result will not change.

For the natural gas pipeline network, the transmission correlation between source nodes and download nodes is important, and the degree of correlation varies with gas transmission characteristics. Assuming that 100 units of natural gas are downloaded from node A, three gas sources (node B, C and D) supply 70 units, 30 units and 0 units respectively. The transmission correlation strength between node B and node A is greater than that between node C and node A, and node D is not related to node A. In other words, the larger the contribution of a source node to a certain download node is, the stronger the transmission correlation between them will be. The gas transmission correlation strength is actually a parameter of flow rate. Therefore, different from conventional statistical correlation concept, the gas transmission correlation strength between source nodes and download nodes is directly defined as the supply volume of each source node to each download node, which can be calculated by the gas flow tracing process in chapter 2. It should be noted that from the perspective of complete hydraulics, this paper only considers the relationship between source nodes and download nodes.

To conduct community detection based on gas transmission characteristics, this paper uses the gas transmission correlation strength to improve the traditional modularity of FGC algorithm and proposes the gas transmission modularity (Q_s). The formula for calculating Q_s is as follows.

$$Q_s=\sum_{ik}\left[\frac{S_{ik}}{N}-\frac{S_i}{N}\frac{S_k}{N}\right]\delta(c_i,c_k) \qquad (3)$$

Here, S_{ik} is the transmission correlation strength between source node i and download node k, which is equal to the supply volume from source node i to download node k. N is the total throughput of pipelines, which is the sum of gas transmission correlation strength between all nodes. N can be expressed as follow.

$$N=\sum_{ik}S_{ik} \qquad (4)$$

S_i and S_k are the total upload volume at source node i and the total download volume at download node k. They can be formulated as follows.

$$S_i=\sum_k S_{ik} \qquad (5)$$

$$S_k = \sum_i S_{ik} \quad (6)$$

The meaning of $\delta(c_i, c_k)$ in Eq.(3) is the same as that of Eq.(1).

The value range of gas transmission modularity is [0,1]. The gas transmission modularity has a clear physical meaning based on gas transmission characteristics, which can be explained as follows. Event *A* is assumed to be the probability that a unit of gas transmission correlation strength starts from node *i*, and event *B* is assumed to be the probability that a unit of gas transmission correlation strength ends into node *k*. When a unit of gas transmission correlation strength is randomly selected, the probability of the gas transmission correlation strength starts from node *i* and end in node *k*, i.e. *P(AB)*, depends on whether the two events are independent or not.

For a gas pipeline network, the gas transmission correlation strength between source nodes and download nodes is determined by the network structure, hydraulic characteristics, flow distribution of source nodes and download nodes. Also, the gas transmission through well designed network is never a random process. Therefore, event *A* and event *B* are not completely independent. The *P(AB)* can be formulated as follows.

$$P(AB) = P(A) \times P(B|A) = (S_i / N) \times S_{ik} / S_i = S_{ik} / N \quad (7)$$

For a random network, event *A* and event *B* are independent events. The *P(AB)* then can be formulated as follows.

$$P(AB) = P(A) \times P(B) = (S_i / N) \times (S_k / N) \quad (8)$$

For the Eq.(3), the first item in brackets corresponds to *P(AB)* of real natural gas pipeline network, and the second item corresponds to *P(AB)* of random network. Therefore, for the given source node *i* and download node *k*, the physical meaning in brackets of Eq.(3) is the distribution probability difference of gas transmission correlation strength between scientifically designed gas network or local network and corresponding random network. $\delta(c_i, c_k)$ represents that the calculation of this distribution probability difference is limited to nodes which are divided into one community. For the gas pipeline network, different community division schemes correspond to different value of Q_s, larger Q_s shows the supply connection between source nodes and download nodes in each community is more close than that in the random network, i.e. the better community detection result based on transmission characteristics.

In summary, this paper replaces the traditional modularity of FGC algorithm with the gas transmission modularity (Q_s) to form a new community detection algorithm. The basic idea of aggregation is adopted and the specific calculation process is as follows: (1) Each node in network is set as a community and then the iterative calculation begins. (2) During each iteration, two communities randomly selected from the network are merged into a new community, and Q_s is calculated. After trying all possible combinations of two communities, the combination with the largest Q_s is taken as the result of this iteration and as the initial working condition of next iteration. The corresponding Q_s value is recorded for final comparison. (3) Repeating step 2 until there is only one community in the network after many mergers, and the iteration ends. (4) The final community division result is the iteration result with the largest Q_s.

4. Case Application

According to the actual operation of a natural gas pipeline network in China, a typical theoretical model is established to analyze the community detection results under three different conditions.

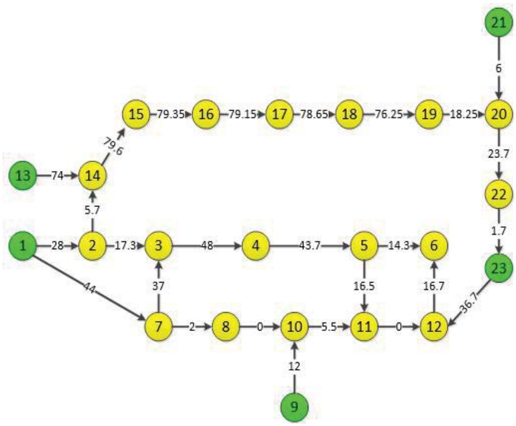

Fig.1. Topological diagram of the theoretical model (Throughput unit: $10^6 m^3/d$).

As shown in Fig.1, the model is mainly composed of 23 nodes. 5 green nodes are source nodes and 18 yellow nodes are download nodes. Among the 5 source nodes, node 1 is a natural gas field, node 9 is an LNG receiving station, node 13 is an interconnection station that receives gas from upstream pipeline, node 21 is a coalbed methane source, node 23 is a gas storage. All the 18 download nodes connect users. For this case, the total upload volume is equal to the total download volume, which is 199×10^6 m^3/d. The daily nomination of each node is listed in Table 1.

Table 1. The daily nomination of each node

Node	Daily Nomination ($10^6m^3/d$)	Node	Daily Nomination ($10^6m^3/d$)
1	72.00	13	74.00
2	5.00	14	0.10
3	6.30	15	0.25
4	4.30	16	0.20
5	12.90	17	0.50
6	31.00	18	2.40
7	5.00	19	58.00
8	2.00	20	0.55
9	12.00	21	6.00
10	6.50	22	22.00
11	22.00	23	35.00
12	20.00	-	-

In Fig.1, the throughput and direction of each edge are labeled. This pipeline transmission scheme is determined by the maximum flow minimum cost algorithm, with the objective of minimizing the total transportation fee under the premise of completing daily nominations as much as possible[15].

The original model condition is called condition 1. The flow tracing algorithm of Bialek is applied and the supply volume of each source node to each download node is obtained under condition 1. The result is shown in Table 2.

Table 2. The gas flow tracing result (condition 1).

Source Node	Down-load Node	Supply Volume ($10^6m^3/d$)	Source Node	Down-load Node	Supply Volume ($10^6m^3/d$)
13	6	0.54	13	12	0.65
13	14	0.09	13	15	0.23
13	16	0.19	13	17	0.46
13	18	2.23	13	19	53.85
13	20	0.38	13	22	15.37

Table 2 (Continued).

Source Node	Down-load Node	Supply Volume ($10^6m^3/d$)	Source Node	Down-load Node	Supply Volume ($10^6m^3/d$)
1	2	5.00	1	3	6.30
1	4	4.30	1	5	12.90
1	6	14.34	1	7	5.00
1	8	2.00	1	11	16.50
1	12	0.05	1	14	0.01
1	15	0.02	1	16	0.01
1	17	0.04	1	18	0.17
1	19	4.15	1	20	0.03
1	22	1.18	9	10	6.50
9	11	5.50	21	6	0.19
21	12	0.23	21	20	0.14
21	22	5.44	23	6	15.93
23	12	19.07	-	-	-

According to the result of gas flow tracing under condition 1, the gas transmission modularity can be determined and the community detection is carried out by the proposed algorithm. The result is shown in Fig.2. Nodes with the same color belong to the same community.

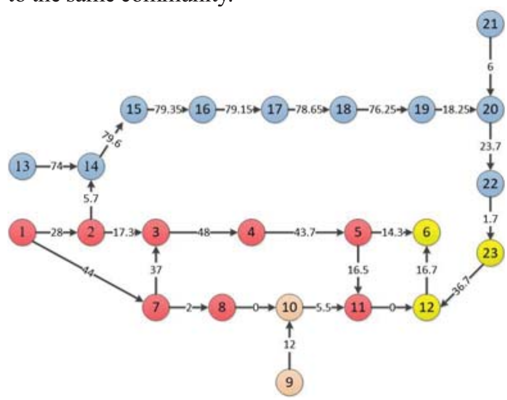

Fig.2. The result of the proposed algorithm under condition 1 (Throughput unit: $10^6m^3/d$).

Seen from Fig.2, the network is divided into 4 communities, and each community is a complete hydraulic system that includes the source node and the download node.

For comparative analysis, the result of FGC algorithm is shown in Fig.3, which also has four communities, but the composition of each community is different.

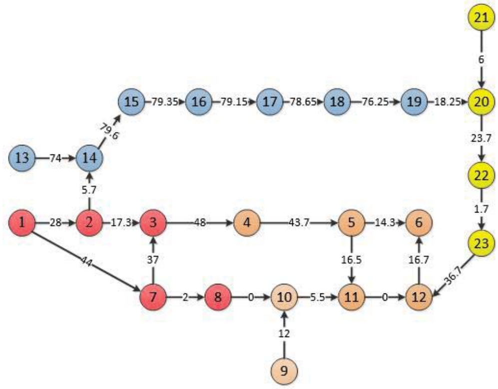

Fig.3. The result of FGC algorithm under condition 1 (Throughput unit: $10^6 m^3/d$).

Condition 2 and condition 3 are set to reflect the frequent change of supply and demand in actual gas pipeline operation. Condition 2 simulates the situation that the shipper reduces the upload volume of LNG for peak shaving when the user demand decreases. Compared with condition 1, the daily nomination of download nodes 11 and 12 decreases by 8×10^6 m^3/d and 4×10^6 m^3/d, respectively. The LNG upload volume at source node 9 decreases from 12×10^6 m^3/d to 0. Condition 3 simulates the situation that the shipper increases the upload volume of LNG for peak shaving when the user demand increases. Compared with condition 1, the daily nomination of download node 22 increases by 13×10^6 m^3/d and the LNG upload volume at source node 9 increases by 13×10^6 m^3/d.

Corresponding community detection results of the proposed algorithm are drawn in Fig.4 and Fig.5 respectively.

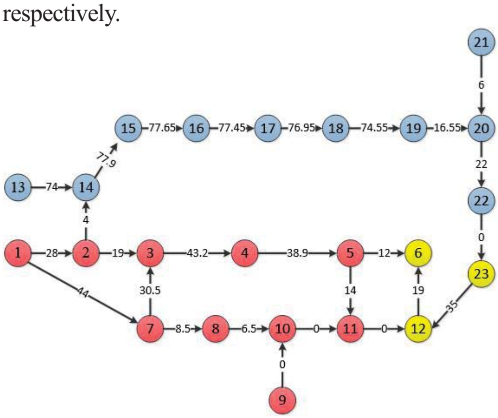

Fig.4. The result of the proposed algorithm under condition 2 (Throughput unit: $10^6 m^3/d$).

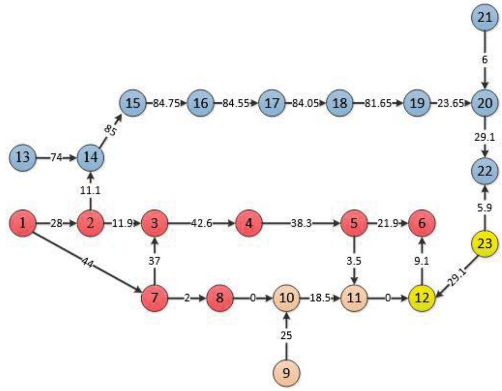

Fig.5. The result of the proposed algorithm under condition 3 (Throughput unit: $10^6 m^3/d$).

Last, the proposed algorithm is applied to an actual gas pipeline network consisting of 110 nodes in China. The community detection result is shown in Fig.6.

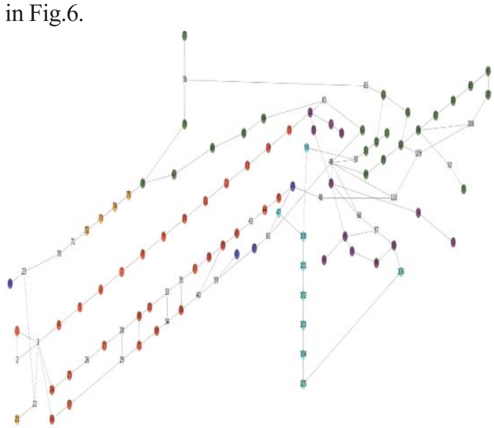

Fig.6. The result of the proposed algorithm for an actual gas pipeline network in China.

5. Case Discussion

The case application mainly focuses on a typical theoretical model and designs three different conditions. Comparing with traditional FGC algorithm, the algorithm proposed in this paper shows three major advantages.

(1) The result of the proposed algorithm is more consistent with the actual gas transmission in natural gas pipeline networks under the same condition.

As can be seen in Fig.3, the FGC algorithm divides source node 9 and download nodes 4, 5, 6, 10, 11 and 12 into the same community under condition 1. However, according to the gas flow tracing result (Table 2), download nodes 4 and 5 are completely supplied by source node 1. Also, 75% of the download volume in download node 11 is supplied by source node 1. And over 95% of the download volume in download node 12 is supplied by source node 23. From the view of gas transmission characteristics, download nodes 4, 5, and 11 have the highest gas transmission correlation strength with source node 1 among all source nodes, and download node 12 has the highest gas transmission correlation strength with source node 23 among all source nodes. They should be correspondingly divided into the same community. By comparison, the result of the proposed algorithm fully matches the gas transmission correlation strength between source nodes and download nodes. As shown in Fig.2, download nodes 4, 5, 11 and source node 1 are divided into the same community, and download node 12 and source node 23 are divided into the same community. Therefore, the proposed algorithm has more reference significance for actual operation.

(2) The result of the proposed algorithm changes adaptively with different operation conditions.

Comparing Fig.2 and Fig.4, the number of communities detected by the proposed algorithm changes from 4 to 3. Due to the LNG upload volume at source node 9 decreases to 0 under condition 2, the download volume of download node 10 supplied by source node 9 under condition 1 are supplemented by the pipeline section from node 8 to node 10 under condition 2. As a result, download node 10 is completely supplied by source node 1. Because the transmission correlation strength between them has changed, the community consisting of nodes 9 and 10 under condition 1 is merged into the adjacent community that controlled by source node 1 under condition 2.

Comparing Fig.2 and Fig.5, due to the increase of LNG upload volume under condition 3, source node 9 supplies additional gas to node 11. The original community consisting of nodes 9 and 10 under condition 1 becomes the community consisting of nodes 9, 10 and 11 under condition 3. In addition, the increased download volume of download node 22 is supplemented by its closest source node 23. As a result, the gas supply from source node 23 to download node 6 decreases, and the decreased amount is supplemented by source node 1. For the download node 6, the gas supply proportion of source node 1 increases from about 46% under condition 1 to about 71% under condition 3. Therefore, download node 6 is assigned to the community controlled by source node 1 under condition 3.

(3) The proposed algorithm can obtain the higher gas transmission modularity.

As previously analyzed, a larger gas transmission modularity (Q_s) corresponds to a better community detection result. By calculation, the Q_s of the proposed algorithm under three different conditions, with value of 0.5424, 0.5599 and 0.5497 respectively, has small difference and maintains at a high level. It shows that the proposed algorithm is stable and corresponding community detection can distinguish the strong and weak gas supply relationships in network. In contrast, the community detection of FGC algorithm keeps the same result under three different conditions. The corresponding Q_s are 0.2662, 0.2475 and 0.2957 respectively.

Seen from Fig.6, an actual gas pipeline network in China is divided into 6 communities and the Q_s is 0.5763. It should be noted that the colorless nodes, which are interconnected stations or valve rooms, are temporarily excluded from community division, because there is no upload gas or download gas from them. If necessary, they can simply be merged into the neighborhood community. Detailed discussions are no longer carried out for confidentiality considerations, but the community detection result is consistent with the actual operating experience, which verifies the effectiveness of the proposed algorithm.

A reasonable community detection is particularly important for the reliability of natural gas pipeline network. First, it will effectively improve the reliability of operation. For the operation of large-scale network, the conventional approach is to use distributed strategy to divide a network into several communities. And the most ideal scenario is to achieve self-supply within community and avoid the gas allocation across different communities as much as possible. Therefore, reasonable community detection can minimize the gas adjustments between communities, reduce the number of flow control by operators, and decrease the number of regulating valves and flow meters, which means the fundamental improvement of operation reliability. By applying the proposed algorithm, the gas throughput is relatively small between communities, and most of gas throughput is distributed within community. Second, it provides a simplified strategy for reliability calculation, greatly improving the computation speed. During the reliability calculation process, large amount of random sampling calculation is required, which means significant computing power is needed to support the large-scale network case. If considering each community as a node and each pipeline between them as an edge, a simplified network can be formed. It will be easier and faster to perform reliability calculation on the simplified network.

6. Conclusion

In this paper, the flow tracing algorithm of Bialek is applied. It can determine the contribution of each source node to each download node and define the gas transmission modularity. Based on gas transmission characteristics, a set of community detection algorithm is then proposed by replacing the traditional modularity of fast greedy community algorithm (FGC algorithm) with the gas transmission modularity. The community detection result with higher value of gas transmission modularity is better, because the greater gas transmission modularity represents the stronger gas transmission correlation between source nodes and download nodes within the community.

The case study shows that the proposed community detection algorithm can dynamically adjust the community division and automatically get the optimal result under different conditions. Each divided community includes the source node, download node and necessary pipeline section, which together form a complete hydraulic system. The calculated gas transmission modularity of the proposed algorithm keeps a high level and stable distribution, which is almost twice the corresponding calculated modularity of FGC algorithm under the same condition. The application result of an actual gas pipeline network in China is consistent with actual operating experience. These all indicate that the proposed community detection algorithm is more helpful for the optimization of gas pipeline network operation.

It should be noted that this paper does not consider the operation linepack that changes continually due to the imbalanced status between supply and demand. It can be solved through the negotiated virtual gas sources and paths, the adjustment of gas storages, etc.

References

[1]GIRVAN M, NEWMAN M E. (2002).Community structure in social and biological networks. Proc Natl Acad USA, 99(12): 7821-7826.
[2]NEWMAN M, GIRVAN M. (2004).Finding and Evaluating Community Structure in Networks. Physical Review E, 69(2): 423-432.
[3]NARDO A D, NATALE M, SANTONASTASO G F, et al. (2013).An Automated Tool for Smart Water Network Partitioning. Water Resources, 27(13): 4493-4508.
[4]NARDO A D, NATALE M, GRECO R, et al. (2014).Ant Algorithm for Smart Water Network Partitioning. Procedia Engineering, 70(1): 525-534.
[5]ZHAO J, JIA R, CHEN L, ZHU T. (2021).Research on fast partition of reactive power and voltage based on deep learning and an improved K-means clustering algorithm. Power System Protection and Control, 49(14): 89-95.
[6]XU Y, MENG Z. (2013).Identify communities in the microblogging based on the GN algorithm. Journal of Guangxi University (Natural Science Edition), 38(06): 1413-1417.
[7]NEWMAN M E. (2006).Modularity and community structure in networks. APS March Meeting. American Physical Society, 8577-8582.
[8]NEWMAN M E. (2004).Fast algorithm for detecting community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys, 69(2): 066133-1-066133-5.
[9]BIALEK J. (1996).Identification of source-sink connections in transmission networks. Fourth International Conference on Power System Control and Management, 200-204.
[10]BIALEK J. (1996).Tracing the flmalingow of electricity. IEE Proceedings-Generation, Transmission Distribution, 143(4): 313-320.
[11]MA L, XIE K, ZHOU, et al. (2004).A survey of power flow tracing algorithm. Journal of Chongqing University, 27(07): 45-49.
[12]KIRSCHEN D, ALLAN R, STRBAC G. (1997).Contributions of individual generators to loads and flows. IEEE Transactions on Power Systems, 12(1): 52-60.
[13]XIE K, LI C, ZHOU N, et al. (2005).A survey of power flow tracing algorithm. Transactions of China Electrotechnical Society, 20(08): 7-11.
[14]WANG X. (2006).Complex network theory and applications. Tsinghua University publishing house, 184 -185.
[15]GUO J, LIU L. (2015). Application of Minimum Cost Maximum Flow Model in Transportation Network Optimization. Modern Business Trade Industry, 36(17): 53-54.