

# Null Model and Community Structure in Heterogeneous Networks <sup>★</sup>

Xuemeng Zhai<sup>1</sup>[0000–0002–3344–3647], Wanlei Zhou<sup>2</sup>, Gaolei Fei<sup>1</sup>, Hangyu Hu<sup>1</sup>,  
Youyang Qu<sup>3</sup>, and Guangmin Hu<sup>1</sup>

<sup>1</sup> University of Electronic Science and Technology of China, Chengdu 611731, China  
hgm@uestc.edu.cn

<sup>2</sup> University of Technology Sydney, Ultimo NSW 2007, Australia

<sup>3</sup> Deakin University, Burwood VIC 3125, Australia

**Abstract.** Finding different types of communities has become a research hot spot in network science. Plenty of the real-world systems containing different types of objects and relationships can be perfectly described as the heterogeneous networks. However, most of the current research on community detection is applied for the homogeneous networks, while there is no effective function to quantify the quality of the community structure in heterogeneous networks. In this paper, we first propose the null model with the same heterogeneous node degree distribution of the original heterogeneous networks. The probability of there being an edge between two nodes is given to build the modularity function of the heterogeneous networks. Based on our modularity function, a fast algorithm of community detection is proposed for the large scale heterogeneous networks. We use the algorithm to detect the communities in the real-world twitter event networks. The experimental results show that our method perform better than other exciting algorithms and demonstrate that the modularity function of the heterogeneous networks is an effective parameter that can be used to quantify the quality of the community structure in heterogeneous networks.

**Keywords:** Heterogeneous Network · Community Detection · Modularity · Twitter Network.

## 1 Introduction

Network science is a fundamental tool to analyze the basic problems of the real-world complex systems, such as social networks, metabolic networks, computer networks and etc [2]. Community detection has become a key research in network science during the past decades [4, 8]. The community refers to the cluster of nodes that are connected densely and community detection focuses on finding the such clusters effectively in the networks. The modularity proposed by Newman based on the null model is the most famous parameter to quantify the quality

---

<sup>★</sup> This work was supported by National Natural Science Foundation of China No. 61571094 and Sichuan Science and Technology Program under Grant 2019YFG0456.

of the community structure in the homogeneous single networks [9]. Based on the modularity function, effective algorithms of the community detection in homogeneous networks are proposed such as the famous BGLL algorithm [1]. The null model and the modularity function are also used in the research on the homogeneous multiplex networks [7, 13]. However, most of the research just focus on the homogeneous networks and there is no effective function to quantify the quality of the community structure in heterogeneous networks.

Most of the real-world networks contain more than one type of the nodes and relationships. For example, in the DBLP networks, there are three types of nodes: authors, papers, and conferences [12] and in twitter event networks, there are two types of nodes: users and events [5]. Such networks are heterogeneous in nature. The community detection method is no longer available for those heterogeneous networks. Therefore, it is necessary to propose the suitable method to detect the communities in heterogeneous networks. The main problem of the heterogeneous community detection is how to deal with the heterogeneous relationships among the different types of the nodes. Researchers propose several method to detect the heterogeneous communities focused on the heterogeneous relationships. However, the basic community structure is ignored so that there is no effective function to quantify the quality of the community structure in heterogeneous networks like the homogeneous modularity function.

In this paper, we propose the null model and modularity function of the heterogeneous networks. The heterogeneous node degree is proposed to replace the node degree of homogeneous networks based on the heterogeneous relationships in the heterogeneous networks. Then we build the null mode of the heterogeneous networks with the same heterogeneous node degree distribution of the original network. The modularity function of the heterogeneous networks is built with the probability of there being an edge between two nodes in the null model. Based on our modularity function, a fast algorithm of heterogeneous community detection is proposed to demonstrate the effectiveness of the modularity. The experimental results show that the community structure of the heterogeneous networks can be exposed effectively through the modularity function. Our findings fill the gap in the field of null model of heterogeneous networks and provide a powerful tool for detecting communities in the complex systems with multiple objects and relationships in many general scientific fields.

The reminder of the paper is structured as follows: The Section II is the related work about our research. The heterogeneous node degree and null model of heterogeneous networks are introduced in Section III. In the Section IV, we discuss the modularity function of the heterogeneous networks and the algorithm of the community detection is shown in Section V. The experiments is presented on Section VI. The Section VII is the conclusions.

## 2 Related Work

The null model in homogeneous networks has the same degree distribution with the original network. The modularity function proposed by Newman basd on the

null model is the most famous parameter that can be used to quantify the quality of the communities in homogeneous network [10]. The modularity refers to the number of edges within communities minus the expected number of such edges in the null model. Based on the modularity, Vincent Blondel et al. [1] propose a fast modularity optimization method called BGLL algorithm. They found the high modularity partitions of large networks in short time and unfolded a complete hierarchical community structure for the network. The method still focused on the homogeneous networks.

Compared with analysis for the homogeneous single networks and multiplex networks, the research on community detection in heterogeneous networks started relatively late. Deng Cai et al. [3] propose a method to find the hidden community in heterogeneous social networks. They built the weighted matrix of different relationships according to the priori community detection results and used the optimized algorithm to calculate the optimal coefficient of each relationship matrix. The coefficient represented the influence of the different relationships on the result of the community detection. The method requires prior knowledge about community detection.

Qiankun Zhao et al. [14] propose a framework of mining different types of communities from web based on the heterogeneity and evolution of web data. They gave the clearly definition of the heterogeneous networks and use a 8-tuple vector to represent them. The features of particular communities were extracted using the PopRank algorithm to build the SVM regression model for the prediction.

Comar et al. [6] use the multi-task learning to classify nodes and detect communities at the same time. They derived two homogeneous subnetworks form a heterogeneous network that contains two types of nodes, one subnetwork for classification and the other for community detection. The author classify the nodes and detect communities through the relevance of the two subnetworks. The methods requires the heterogeneous networks must be bipartite.

Qiu et al. [11] focus on the overlapping community detection of the heterogeneous social networks. They propose an algorithm called OcdRank (Overlapping Community Detection and Ranking) combining the overlapping community detection and community-member ranking together in directed heterogeneous social networks. The algorithm still works on bi-type heterogeneous social networks.

Our work differs from those found on the literature because the null model and modularity are the basic theory of the community detection in homogeneous networks. We propose the two basic conceptions of the heterogeneous networks and focus on the community structure itself with the considering of the heterogeneity in heterogeneous networks. The work is original and unprecedented.

### 3 Null Model of Heterogeneous Networks

#### 3.1 Heterogeneous Networks and Heterogeneous Node Degree

We first introduce the basic conception of the heterogeneous networks. In this paper, we use the set of adjacency matrices to describe a heterogeneous networks ( $HW$ ) as  $HW = \{A^S, \dots, H^{SR}, \dots\}, S, R \in T$ , where  $T$  refers to the type of nodes.  $A^S = (a_{ij}^S)_{N_S \times N_S}$  donated as the adjacency matrix of the same-type nodes, where  $N_S$  refers to the number of  $S$ -type nodes.  $H^{SR} = (h_{ij}^{SR})_{N_S \times N_R}$  donated as the adjacency matrix of two different types of nodes. In the representation, we just separate the homogeneous nodes and heterogeneous nodes to ensure importance of the heterogeneous links in the community detection of the heterogeneous networks.

The existing null model of the homogeneous single network is proposed by Newman and has the same distribution of the node degree with the original network. To build the null model of the heterogeneous networks, we should first propose a new parameter to describe the basic connection among the different types of nodes in the heterogeneous networks like the node degree in the homogeneous networks. Therefore, we first define the heterogeneous node degree as follows:

**Definition 1 Heterogeneous Node Degree:** For each type of nodes, there are neighbors of the  $S$ -type node  $i$ . The heterogeneous node degree refers to the number of neighbors of different types from a node  $i$ . We give the  $u_i^{SR}$  to represent the heterogeneous node degree of types  $R$  for the  $S$ -type node  $i$ . When  $S = R$ , the heterogeneous node degree  $u_i^{SS}$  becomes the homogeneous node degree  $k_i$ .

Therefore, the node degree in the heterogeneous networks is divided into two parts: the homogeneous node degree  $k_i$  and the heterogeneous node degree of all types  $\sum_R u_i^{SR} (S \neq R)$ .

#### 3.2 Null Model of the Heterogeneous Networks

With both homogeneous and heterogeneous node degree, we give definition of the null model of the heterogeneous networks:

**Definition 2 Null Model of Heterogeneous Networks:** The null model of the heterogeneous networks refers to those network models that has the same set of types of nodes  $T$ , number of homogeneous nodes  $N$ , number of heterogeneous  $U$ , distribution of homogeneous node degree  $P(k)$  and distribution of heterogeneous node degree  $P(u)$  with the original network, while otherwise is taken to be an instance of the random network.

For each two types of the nodes, there is a distribution of heterogeneous node degree. Therefore, there are  $|T|^2 - |T|$  distribution of heterogeneous node degree in a heterogeneous network, where  $T$  refers to the set of types of nodes.

#### 3.3 Random Walk on Heterogeneous Networks

Here we use the random walk theory to build the null model of the heterogeneous networks. The process can be explained by the Laplacian Dynamics. Considering

a homogeneous network, if there is an edge between node  $i$  and node  $j$ , the two nodes are regarded as reachable. We suppose that there is a walker walking randomly among the nodes in the networks and each walk from one node to the other is completely independent and random. The process is actually a Markov process in which each walk has no relationship with the last time. Therefore, the probability of the walker walking from a arbitrary node  $j$  to node  $i$  and staying at node  $i$ ,  $\dot{p}_i$  is:

$$\dot{p}_i = \sum_j \frac{a_{ij}}{k_j} p_j - p_i \quad (1)$$

where  $p_j$  refers to the probability of the walker staying at the node  $j$ . Differently, in a heterogeneous network, the edges among nodes is divided into homogeneous edges (edges between two same-type nodes) and heterogeneous edges (edges between two different-type nodes). Therefore, when the walker walks in the heterogeneous network, both homogeneous and heterogeneous edges should be considered. The probability of the walker walking from a arbitrary  $R$ -type node  $j$  to  $S$ -type node  $i$  and staying at node  $i$ ,  $\dot{p}_i^S$  is:

$$\dot{p}_i^S = \sum_{j,R} \frac{a_{ij}^S \delta_{SR} + h_{ij}^{SR} \bar{\delta}_{SR}}{\kappa_j^R} p_j^R - p_i^S \quad (2)$$

where  $a_{ij}^S$  refers to the connection relationship between the two nodes  $i$  and  $j$  that belong to the same type  $S$ ;  $h_{ij}^{SR}$  refers to the connection relationship between  $S$ -type node  $i$  and  $R$ -type node  $j$ ;  $\delta_{SR}$  is the reaction function; When  $S = R$ ,  $\delta_{SR} = 1$ ; When  $S \neq R$ ,  $\bar{\delta}_{SR} = 1$ ;  $p_j^R$  refers to the probability of the walker staying at the  $R$ -type node  $j$ ;  $\kappa_j^R$  refers to total degree of the  $R$ -type node  $j$ , that is the sum of homogeneous degree and heterogeneous degree, donated as:

$$\kappa_j^R = k_j^R + \sum_S u_j^{RS} \quad (3)$$

Therefore, we give the conditional probability of the walker walking from  $R$ -type node  $j$  to  $S$ -type node  $i$  of in the null model of heterogeneous networks, donated as:

$$p(i|j^R) = \frac{k_j^R}{\kappa_j^R} \frac{k_i^S}{2M_S} \delta_{SR} + \frac{u_j^{RS}}{\kappa_j^R} \frac{u_i^{SR}}{2M_{SR}} \bar{\delta}_{SR} \quad (4)$$

where  $M_S$  refers to the edge number among the  $S$ -type nodes and  $M_{SR}$  refers to the edge number between  $S$ -type nodes and  $R$ -type nodes. When the Markov process of random walk reaches steady state, the steady probability of the walks staying at the  $R$ -type node  $j$  is donated as:

$$p_j^{R*} = \frac{\kappa_j^R}{2M} \quad (5)$$

where  $M$  refers to the total number of the edges in the heterogeneous network. Therefore, the joint probability of the walker walking from  $R$ -type node  $j$  to

$S$ -type node  $i$  in the null model is:

$$\begin{aligned}
 p(Si, Rj) &= p_i^S | j^R \times p_j^{R*} \\
 &= \left( \frac{k_j^R}{\kappa_j^R} \frac{k_i^S}{2M_S} \delta_{SR} + \frac{u_j^{RS}}{\kappa_j^R} \frac{u_i^{SR}}{2M_{SR}} \bar{\delta}_{SR} \right) \frac{\kappa_j^R}{2M} \\
 &= \frac{1}{2M} \left( \frac{k_j^R k_i^S}{2M_S} \delta_{SR} + \frac{u_j^{RS} u_i^{SR}}{2M_{SR}} \bar{\delta}_{SR} \right)
 \end{aligned} \tag{6}$$

The  $p(Si, Rj)$  is the probability of there being an edge between  $S$ -type node  $i$  and  $R$ -type node  $j$  in the null model of the heterogeneous network. The equation is divided into two parts: the homogeneous part and the heterogeneous part. The homogeneous part is the same with the probability in the homogeneous null model and the the homogeneous part represents the heterogeneous relationships in the heterogeneous network. With this edge-building probability, we could build the modularity function of the heterogeneous network based on the null model.

#### 4 Modularity Function of Heterogeneous Networks

The modularity function is first proposed by Newman in 2006. The modularity  $Q$  = (the number of edges within communities - the expected number of such edges in the null model). The null model here is homogeneous and the modularity proposed by Newman is still built for the homogeneous networks. Similarly, when we replace the null model of homogeneous networks by the one of heterogeneous networks, we can build the modularity function of the heterogeneous networks. Here, we give the definition of the modularity function of heterogeneous networks:

**Definition 3** Modularity Function of Heterogeneous Networks: The modularity function of heterogeneous networks  $Q_h$  = (the number of edges within communities in heterogeneous networks - the expected number of such edges in the heterogeneous null model) and normalized by the total degree of the networks:

$$Q_h = \frac{1}{2M} \sum_{ijSR} [E(Si, Rj) - P(Si, Rj)] \delta(g_{Si}, g_{Rj}) \tag{7}$$

where  $E(Si, Rj)$  refers to the number of edges within communities in heterogeneous networks and  $P(Si, Rj)$  refers to the expected number of such edges in the heterogeneous null model.  $\delta(g_{Si}, g_{Rj}) = 1$  if the  $S$ -type node  $i$  and  $R$ -type node  $j$  belong to the same community, otherwise  $\delta(g_{Si}, g_{Rj}) = 0$ . In the equation 6, we obtain the probability of there being an edge between  $S$ -type node  $i$  and  $R$ -type node  $j$  in the null model of the heterogeneous network. Therefore, the  $P(Si, Rj)$  is donated as:

$$\begin{aligned}
 P(Si, Rj) &= p(Si, Rj) * 2M \\
 &= \frac{k_j^R k_i^S}{2M_S} \delta_{SR} + \frac{u_j^{RS} u_i^{SR}}{2M_{SR}} \bar{\delta}_{SR}
 \end{aligned} \tag{8}$$

The actual number of edges between two nodes in heterogeneous networks can be represented as:

$$E(Si, Rj) = A_{ij}^R \delta_{SR} + H_{ij}^{SR} \bar{\delta}_{SR} \quad (9)$$

With the equation of the modularity function in details:

$$Q_h = \frac{1}{2M} \sum_{ijSR} [(A_{ij}^S - \frac{k_j^R k_i^S}{2M_S}) \delta_{SR} + (H_{ij}^{SR} - \frac{u_j^{RS} u_i^{SR}}{2M_{SR}}) \bar{\delta}_{SR}] \delta(g_{Si}, g_{Rj}) \quad (10)$$

In the equation 10, we divide the modularity function of heterogeneous networks  $Q_h$  into two parts, the homogeneous part  $A_{ij}^S - \frac{k_j^R k_i^S}{2M_S}$  and the heterogeneous part  $H_{ij}^{SR} - \frac{u_j^{RS} u_i^{SR}}{2M_{SR}}$ . Therefore, the modularity can be understood as the sum of both homogeneous and heterogeneous part, which reflects the whole kinds of relationships in the heterogeneous networks.

## 5 Community Structure and the Fast Algorithm in Heterogeneous Networks

Similar with the homogeneous networks, there are also community structure in the heterogeneous networks, that is, the set of multi-type nodes that are connected closely. The modularity of heterogeneous networks can be used to quantify the quality of the heterogeneous community structure. When the modularity gets max, the results of the community detection are the best.

We start from the basic structure of the networks to detect the heterogeneous communities. Therefore, we do not distinguish the type of nodes when detecting the communities. Which community a node belongs to is decided by the change of modularity function when it joins the community. The final results of each heterogeneous community will contain at least one type of nodes or more. It all depends on the maximum modularity function of the heterogeneous networks. After the community detection, we could extract the same-type nodes in each community to get the homogeneous node clusters.

Based on the modularity function of heterogeneous networks, we give a fast algorithm to detect the communities in the heterogeneous networks. The process of the algorithm is shown in algorithm 1. The time complexity of algorithm 1 (FAHCD) is  $O(N \times \max(\kappa_i))$ . The algorithm is based on the famous fast algorithm BGLL of the homogeneous networks. In the large networks, the  $\max(\kappa_i)$  is far less than the number of nodes  $N$ . Therefore, the time complexities of the algorithm 1 is close to  $O(N)$ .

## 6 Experiments

We use the FAHCD to detect the heterogeneous communities of the twitter event networks we build through the real-world data. The twitter data is collected from

---

**Algorithm 1** Fast Algorithm of Heterogeneous Community Detection (FAHCD).

---

**Require:** The adjacency matrix set of heterogeneous network,  $HW = \{A^S, \dots, H^{SR}, \dots\}, S, R \in T$

**Ensure:** The results of the heterogeneous communities,  $C_h^K = \{C_{h_1}, C_{h_2}, \dots, C_{h_n}\};$

- 1: initial  $HW^0 = HW;$
  - 2: **repeat**
  - 3:   Regarding each node in  $HW^k$  as a community initially.  $C_h^k = \{Node_1, Node_2, \dots, Node_N\}$ , where  $N$  is the total number of nodes in  $HW^k;$
  - 4:   Computing the increment of modularity  $\Delta Q_h^{ij}$  between each node  $i$  and its each neighbor  $j$  in the heterogeneous network;
  - 5:    $\Delta Q_h^{ij} = \frac{1}{2M} \{ \sum_{z \in g_j} [(A_{iz}^S - \frac{k_i^S k_z^R}{2M_S}) \delta_{SR} + (H_{iz}^{SR} - \frac{u_i^{SR} u_z^{RS}}{2M_{SR}}) \bar{\delta}_{SR}] - \sum_{z \in g_i} [(A_{iz}^S - \frac{k_i^S k_z^R}{2M_S}) \delta_{SR} + (H_{iz}^{SR} - \frac{u_i^{SR} u_z^{RS}}{2M_{SR}}) \bar{\delta}_{SR}] \};$
  - 6:   For the node  $j$  with the max  $\Delta Q_h^{ij}$  with node  $i$ , adding the node  $i$  into the community with node  $j$ ;
  - 7:   Updating the set of communities  $C_h^k$  after the aggregation of  $C_h^{k-1}$  in step 6;
  - 8:   Regarding each type of nodes in the new community in  $C_h^k$  as the new specific-type node; Regarding connections among the nodes as the self-loop of the new node with the weight of number of connections; Regarding edges between two different type of nodes as the new edge between two new nodes with the weight of number of edges; Generating a new heterogeneous network  $HW^k$
  - 9: **until**  $\Delta Q_h < 0$  of all nodes;
  - 10: **return**  $C_h^K;$
-

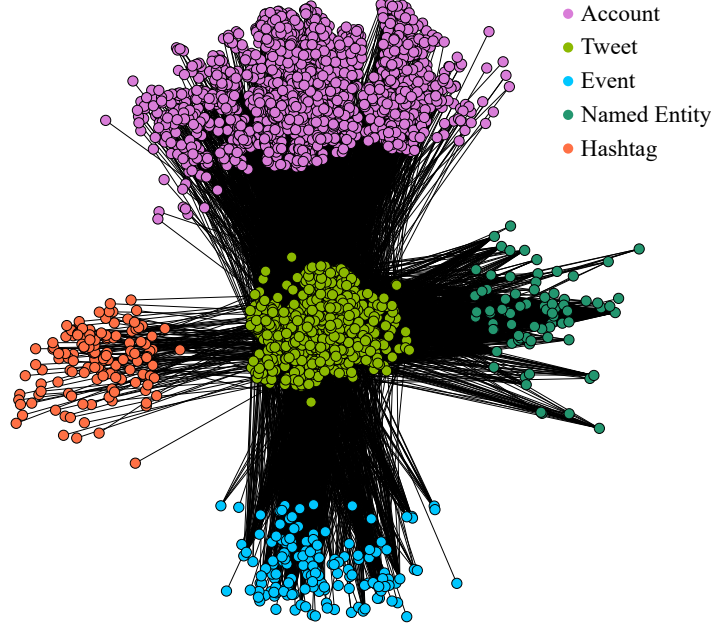
the Twitter API. The MongoDB database is used to store the collected data. After pre-processing, including tweet language filtering, spam tweet filtering, useless field filtering and text content filtering, we obtained the valuable tweets and accounts. The Named Entity Recognition (NER) is used to extract the name of related people in each tweet and to extract the hashtag by the key symbol #. The twitter events are clustered based on the text similarity among the tweets. We cluster a large number of tweets with high text similarity to detect a twitter event that occur in the Twitter space. Then we build the twitter event networks with 5 type of nodes: Account, Tweet, Event, NameEntity and Hashtag. We capture 4 type of relationships among the 5 type of nodes. The networking rules are shown in table 1

**Table 1.** The networking rules of the Twitter event network

| Type | Name                 | Description                                       |
|------|----------------------|---------------------------------------------------|
| Node | Account              | The twitter accounts of the users.                |
| Node | Tweet                | The short message written by the twitter users.   |
| Node | Event                | The events detected in the Twitter space.         |
| Node | Named Entity         | The name of related people detected in tweets.    |
| Node | Hashtag              | The content tag for the tweets.                   |
| Edge | Account and Tweet    | Connected if the tweet is written by the account. |
| Edge | Tweet and Event      | Connected if the tweet belongs to the event.      |
| Edge | Tweet and NameEntity | Connected if the name appears in the tweet.       |
| Edge | Tweet and Hashtag    | Connected if the tweet has the hashtag.           |

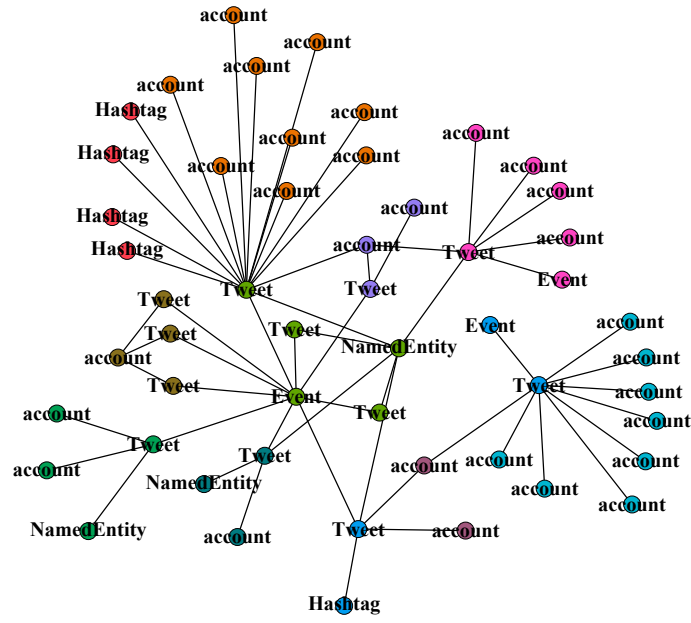
We collected Twitter data about the UK elections from May 12nd, 2017 to June 10th, 2017. The Twitter event network in 30 days we built consisted of 70,536 account nodes, 32,593 tweet nodes, 2,618 event nodes, 1745 named entity nodes, and 1462 hashtag nodes. The twitter event network on May 12nd, 2017 is visualized in figure 1. There are 3,459 nodes and 4,329 edges including 2232 account nodes, 854 tweet nodes, 161 event nodes, 134 named entity nodes and 78 hashtag nodes. As shown in figure 1, the core-type of the nodes are the

tweet nodes. They are connected with the rest other type of nodes to form the 4 types of the edges in the Twitter event networks. The rest 4 types of nodes are disconnected.



**Fig. 1.** The Twitter Event Network about UK Elections.

We detected the heterogeneous communities on the Twitter event networks we built in 30 days using FAHCD. The partial results of the community detection are visualized in figure 2. Nodes in the same color belong to the same communities. For visualization, we delete lots of nodes and edges of the network. The results show that different types of nodes could be divided into the same community because the dense connection among them such as the green nodes in the center in figure 2. Because of the different types of connection, different types of nodes could be divided into different communities, such as the orange nodes and the red nodes in the left top of figure 2. The orange nodes are account and the red nodes are hashtag. They are all connected with the tweet nodes in green but they are divided into different communities just because the connections among them are heterogeneous. Therefore, our algorithm detects the communities using modularity based on the heterogeneous structure itself of the heterogeneous networks. It can not be replaced by transferring the heterogeneous networks into homogeneous networks and using the homogeneous community detection methods, which ignores the critical heterogeneous structure information.



**Fig. 2.** The Partial Results of the Heterogeneous Community Detection on Twitter Event Network.

The 504 communities are detected in the Twitter event network consists of 108,954 nodes. 84% communities contains the whole 5 types of nodes and only 4 communities contains just 2 types of the nodes. Such communities are small with less than 100 nodes and made up by the account nodes and tweet nodes. They are not connected with any other types of nodes. We manually labeled the election position of 1350 account nodes as the ground-truth to quantify the performance of our community detection. The results are shown in table 2. Here we got three three position about the UK election: proposition, neutral and opposition. The position of a community is determined by the position of most of its nodes. If the most of nodes are proposition in a community, all of nodes in the community are regarded as the proposition node. Therefore, we could calculate the accuracy as follows:

$$Accuracy = \frac{N_{correct}}{N_{total}}, \quad (11)$$

where  $N_{correct}$  refers to the number of nodes with the correct position and  $N_{total}$  refers to the total number of nodes in the communities of a same position. From the results, we could conclude that people may communicate with each other who has the same position in Twitter. Our algorithm detect the cluster of most people with the same position on the UK election from the Twitter event heterogeneous networks. The error less than 10% is caused by those active nodes and some junk accounts who may connect with people of any position.

**Table 2.** The Performance of FAHCD Based on Ground-truth

| Election Position | Number of Communities | Accuracy |
|-------------------|-----------------------|----------|
| Proposition       | 256                   | 92.3%.   |
| Neutral           | 127                   | 94.7%.   |
| Opposition        | 121                   | 91.2%.   |

## 7 Conclusion

The results in the experiment section demonstrate the advantageous heterogeneous community detection performance on real-world Twitter event networks based on the null model of heterogeneous networks. Our method could deal with large-scale heterogeneous networks on a almost linear time complexity. Based on the FAHCD, we find the cluster of most people with the same position on the UK election from the Twitter event heterogeneous networks we built. The accuracy of all three position is over 90%, which show a great performance of our method on heterogeneous community detection. The community we detected contains more than one type of nodes based on the structure of heterogeneous networks and could be further divided into several homogeneous communities based on the type of each node.

In a general sense, the null model of heterogeneous networks is a general null model for any systems with multi-type of nodes including social networks. The rationality of the model can be explained by the traditional random-walk theory. The general significance of the model is that in addition to heterogeneous community structure, many other specific properties of heterogeneous networks can be revealed through the model. These properties, including motif identities, propagation-rate threshold, redundancy-distribution correlations and synchronization-state stability, have already been shown to be important in homogeneous network research. Additionally, the null model of heterogeneous networks can be used in directed networks based on in-and-out heterogeneous degree. Our future work is based on such extensions of our null model and its high-order representations, which may lead to some problems involving the applications of all systems with multi-type of nodes that can be described by heterogeneous networks.

Finally, the null model of heterogeneous networks provides a powerful tool for the structure analysis of complex systems with multi-type of nodes. Through comparisons, the specific nature of these systems can be exposed quantitatively by the model. We believe that the null model of heterogeneous networks can give rise to much stronger and more general applications in many areas, including social science, Internet topology, bioscience, engineering, economics, and education, where systems can be described by heterogeneous networks. To accomplish this, much more work needs to be done to gain a deeper understanding of the model and its high-order representations. We hope that many more attributes of the complex systems can be modelled and analysed through the null model of heterogeneous networks.

## Acknowledgment

This work was supported by National Natural Science Foundation of China No. 61571094 and Sichuan Science and Technology Program under Grant 2019YFG0456. The data sets used to obtain the results in this manuscript are collected through Twitter API (<https://dev.twitter.com/>).

## References

1. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of community hierarchies in large networks. *J Stat Mech* **abs/0803.0476** (2008)
2. Börner, K., Sanyal, S., Vespignani, A.: Network science. *Annual review of information science and technology* **41**(1), 537–607 (2007)
3. Cai, D., Shao, Z., He, X., Yan, X., Han, J.: Mining hidden community in heterogeneous social networks. In: *Proceedings of the 3rd international workshop on Link discovery*. pp. 58–65. ACM (2005)
4. Cao, Y., Zhang, G., Li, D., Wang, L.: Online energy management for smart communities with heterogeneous demands. In: *2018 IEEE Global Communications Conference (GLOBECOM)*. pp. 1–6. IEEE (2018)

5. Chen, F., Neill, D.B.: Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1166–1175. ACM (2014)
6. Comar, P.M., Tan, P.N., Jain, A.K.: Simultaneous classification and community detection on heterogeneous network data. *Data mining and knowledge discovery* **25**(3), 420–449 (2012)
7. Mucha, P.J., Richardson, T., Macon, K., Porter, M.A., Onnela, J.P.: Community structure in time-dependent, multiscale, and multiplex networks. *science* **328**(5980), 876–878 (2010)
8. Newman, M.E.: The structure and function of complex networks. *SIAM review* **45**(2), 167–256 (2003)
9. Newman, M.E., Girvan, M.: Finding and evaluating community structure in networks. *Physical review E* **69**(2), 026113 (2004)
10. Newman, M.E., Strogatz, S.H., Watts, D.J.: Random graphs with arbitrary degree distributions and their applications. *Physical review E* **64**(2), 026118 (2001)
11. Qiu, C., Chen, W., Wang, T., Lei, K.: Overlapping community detection in directed heterogeneous social network. In: International Conference on Web-Age Information Management. pp. 490–493. Springer (2015)
12. Sun, Y., Yu, Y., Han, J.: Ranking-based clustering of heterogeneous information networks with star network schema. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 797–806. ACM (2009)
13. Zhai, X., Zhou, W., Fei, G., Liu, W., Xu, Z., Jiao, C., Cai, L., Hu, G.: Null model and community structure in multiplex networks. *Scientific Reports* **8**(1), 3245 (2018)
14. Zhao, Q., Bhowmick, S.S., Zheng, X., Yi, K.: Characterizing and predicting community members from evolutionary and heterogeneous networks. In: Proceedings of the 17th ACM conference on Information and knowledge management. pp. 309–318. ACM (2008)