Future Generation Computer Systems 107 (2020) 458–468

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Exploring the transition behavior of nodes in temporal networks based on dynamic community detection

Tianpeng Li [a], Wenjun Wang [a], Xunxun Wu [a], Huaming Wu [c], Pengfei Jiao [b,a,*], Yandong Yu [a,*]

[a] College of Intelligence and Computing, Tianjin University, Tianjin, China
[b] Center of Biosafety Research and Strategy, Tianjin University, Tianjin, China
[c] Center of Applied Mathematics, Tianjin University, Tianjin, China

ARTICLE INFO

Article history:
Received 7 October 2019
Received in revised form 27 December 2019
Accepted 2 February 2020
Available online 13 February 2020

Keywords:
Dynamic complex network
Node transition
Node-level feature
Binary classification
Average neighbor degree

ABSTRACT

Community detection and community evolution tracking are two important tasks in dynamic complex network analysis. Recently, a variety of models and methods have been proposed for detecting the community structure and analyzing their evolution. However, all these methods are only committed to improving the performance of community detection or identifying evolutionary events, ignoring the internal relevance between the structure of each snapshot of the dynamic network and the evolution pattern of communities, especially the structural features of nodes and their dynamic transition behavior. To cope with this problem, we firstly conduct experiments on 15 real-world dynamic networks to explore the transition behavior of nodes in dynamic networks, which is one of the most influential evolutionary patterns in temporal community detection. Firstly, we obtain the temporal community structure based on very successful temporal community detection methods. Secondly, we extract features of nodes based on the structure of the dynamic network, and take the community transition behavior of nodes as the binary classification problem. Finally, we use the decision tree to find the node-level features that have a general impact on node transition. Experiments indicate that the degree and average neighbor degree of nodes have the most common indispensable impact on the node transition behavior, which are very helpful for modeling dynamic complex networks in future.

1. Introduction

Complex network analysis [1,2] has received increasing attention from researchers in different fields, including computer science, social science, and physical science [3–5]. Complex networks always consist of nodes and edges, which represent the objects and the interactions between the objects, respectively. For example, in a social network, nodes could be the social accounts and edges represent the following or followed relationships between accounts. As one of the most important and powerful data structures, analyzing and modeling complex networks can be used for many missions, such as social interaction pattern analysis, social recommendation and protein functional modules recognition. As the most fundamental tasks in complex networks, node identification, link prediction and information dissemination have been widely studied and concerned. In addition, community detection is also one of the most significant tasks, which is usually defined as identifying tightly linked subgraphs from complex networks and benefiting from other tasks.

In general, detecting community structures can help us recognize meaningful modules of a network. A variety of works for community detection have been developed, such as modularity-based methods [6], model-based methods [7,8] and random walk-based methods [9–11], where comprehensive surveys can be seen in [12,13]. However, all these methods assume that the target network is static, that is, the network structure is invariant. Virtually, the network structure varies over time, i.e. dynamic networks. More specifically, in a dynamic network, the nodes may birth or death with time and links between two nodes may appear or disappear. For dynamic network modeling, we usually reply to it as a series of snapshots or slices, each of which can be regarded as a static network. From the perspective of community detection, compared with static networks, detecting the dynamic community poses new challenges [14], among which, how to fuse consecutive snapshot networks to improve performance of community detection and how to describe the evolution of communities are the most important.

Take a co-author network as an example, just as shown in Fig. 1, we show two snapshots of the dynamic network based on the DBLP data [15]. The nodes and edges are the authors and their cooperative relationship, and nodes with the same color

* Corresponding authors.
E-mail addresses: pjiao@tju.edu.cn (P. Jiao), cfssyyd@163.com (Y. Yu).

https://doi.org/10.1016/j.future.2020.02.006
0167-739X/

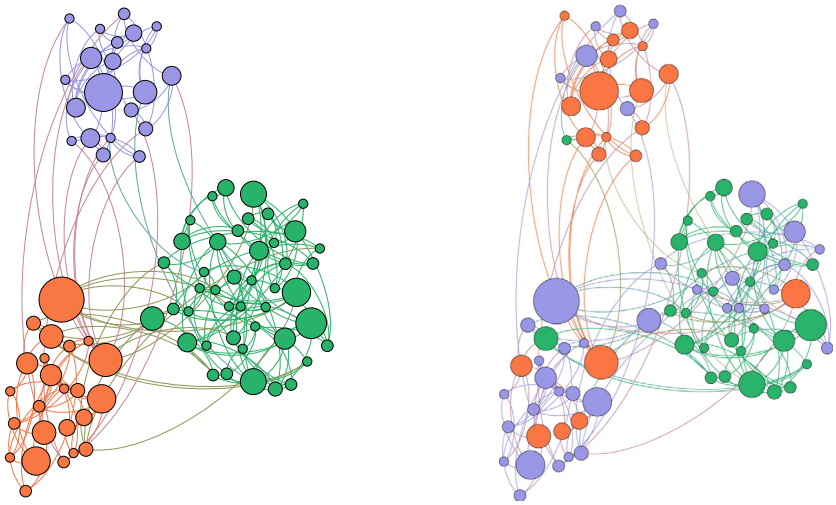


Fig. 1. A co-author network with two snapshots. The left figure is the co-authorship of three communities in the previous snapshot, corresponding to the data mining (green), database (blue) and machine learning (orange), while the right figure shows the community assignment changes in the next snapshot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

represent the same community to which they belong. These three communities are the authors from data mining, database and machine learning, respectively. From the last snapshot to the next, a very important phenomenon is that the research field of some nodes has changed, for example, an author from the database joins into the data mining with the time going by and varying of the network. This is a critical behavior of community detection in dynamic networks, i.e. the transition behavior of nodes, which is the most widely considered dynamic pattern and also is our concern in this paper.

In recent years, more and more attention has been paid to dynamic community detection and different methods have been proposed, including two-step methods, evolutionary clustering methods and model-based methods. Two-step based methods [16,17] usually apply a static community detection algorithm to each snapshot, and then perform community matching step at adjacent time slices. This kind of methods is not accurate enough because data in the real world is often noisy. Moreover, such a two-step process usually results in unstable community structures and consequentially, unwarranted community evolution [15]. Evolutionary clustering is firstly devoted to clustering the stream data and has been developed for dynamic community detection, the previous or historical network or community information are integrated into the community detection in following subsequent network snapshots, such as the evolutionary spectral clustering, dynamic non-negative matrix factorization and multi-objective evolutionary clustering [18,19], this type of methods is still the most widely studied and used. The model-based methods [20,21] usually define a series of network generation mechanisms to reconstruct the dynamic complex network and analyze the evolution of communities, such as the dynamic stochastic block model DSBM [15], which denoted the dynamic pattern based on the classic SBM and transforming community detection and evolution into the parameter estimation. On the whole, the model-based methods have very high computational complexity.

As we all know, all the existing methods for dynamic community detection are focusing on the performance of community detection and the evolutionary patterns or events, while ignoring the internal relevance between the structure varying of dynamic network and the evolution pattern of communities. Therefore, we are interested in, how the structural information of nodes affects the community transitions. In other words, community evolution is usually driven by node transition, and the relationship between the transition behavior and the local varying of nodes is our concern. Although some model-based methods (e.g. [22,23]) use the degree of nodes to improve the accuracy of community detection, these methods only make the node distribution within a community following the power law and do not reveal the relationship between nodes degree and community evolution. As we have discussed, what kind of nodes are more likely to transfer their communities? Are there more statistical features related to the transfer behavior of nodes? Which is the most important feature? We believe that this could help us design more suitable models for community discovery in dynamic networks.

Our motivation is to explore which local structure information or features of the node has important impact on the transition behavior of nodes in dynamic networks, and which structural feature has a larger influence and which one has a small impact. So in this paper, for a given dynamic network, we firstly obtain its community structure based on three very successful temporal community detection methods. Then, we extract the ten features of nodes based on the structure of the previous snapshot network, and take the community transition behavior of nodes as the binary classification problem. In detail, we use the decision tree as the classification model to find the node-level features that have a general impact on node transition and analyze the community evolution on all the snapshots of the dynamic network. We take the framework on 15 real-world dynamic networks shows that the degree and average neighbor degree of nodes are the most two important features impacting on the node transition behavior. We believe that this is very helpful for modeling dynamic complex networks in future. The specific contributions of this paper are as follows:

- As far as we know, this paper is the first exploration of the problem that what kind of nodes is more likely to transfer its community, it is the most important behavior in dynamic networks.
- We extract the community features of the nodes belonging and features of the nodes themselves, and treat the node's community transition as a binary classification problem, then use these features to classify whether the nodes are transferred or not.
- We find that the important common feature of the node's community transition is node's average neighbor degree and node's degree. And node's average neighbor degree is even more important than node's degree, which is inconsistent with our previous understanding.

2. Related work

Community detection is a fundamental task in complex network analysis, which can offer insight into the network formation mechanism and prediction [13,24].

There have been a variety of methods proposed for community detection, including modularity optimization methods, spectral clustering methods and model-based methods. For example, Liu et al. [25] proposed a modularity optimization method using simulated annealing with a k-means iterative procedure to realize the model selection, which outperforms most of the similarity methods. Some other methods [26] detect clusters of networks by utilizing the spectral properties of the graph, but when the network is sparse, the eigenvalues of the community-related eigenvectors are not disparate, which may make spectral clustering unstable. Krzakala et al. [27] proposed a spectral algorithm based on a non-backtracking walk to solve this problem on directed networks. Karrer et al. [7] proposed a model-based method called the degree corrected stochastic block model, in which nodes in the same community can have heterogeneous degrees. That is in line with real-world data. The detailed review can be seen in [13]. However, all these methods are only designed for static networks without considering the temporal information.

Dynamic community detection needs to solve two key subproblems. One is detecting community structure of each snapshot, and the other is matching communities across consecutive time slices or tracking community evolution. Most previous studies have addressed these two issues separately.

For the first problem, previous works can be divided into two-step methods, evolutionary clustering and model-based methods. The two-step approaches solve this problem by performing a static community detection method on each snapshot and then matching communities between consecutive snapshots. Tajeuna et al. [28] proposed a two-step method, which uses a similarity measure that involves the global temporal aspect of the network under investigation to match the communities in different time slices. Evolutionary clustering introduces the community division information from previous snapshots when performing the community detection approach to the current snapshot [19]. TILES [29] is a state-of-the-art evolutionary clustering method, which dynamically recomputes community membership of nodes whenever a new interaction takes place. This strategy makes TILES fit for large networks and its accuracy is higher than most existing algorithms. Model-based methods, like dynamic stochastic block model [15], considering the dynamic network from the perspective of generating model, the mechanism of the network is constructed and the community structures are obtained by parameter estimation [30,31].

Meanwhile, for community evolution, this exciting work [32] summarizes community evolution into identifying some events, and then uses these events to carry out community and node evolution behavior. Palla et al. [16] is the first to give the definition of six community evolution events, including birth, death, merging, splitting, growth and contraction. They first used a clique percolation method to detect communities in each snapshot, then matched community evolution events and analyzed community evolutionary and node behavior prediction in consecutive snapshots by defining an auto-correlation function. Greene et al. [33] proposed a standard dynamic network data set based on community evolution events, which has been widely used in dynamic community detection. Asur et al. [34] not only define five community events, but also define four node-level events, including appear, disappear, join and leave, to capture the influence of the behavior of nodes on communities. But the dynamic behaviors of different nodes are complex, and they have different effects on the community. However, these works did not take into account that the evolutionary behaviors of different nodes are distinguishing, and the impact on the community is also different.

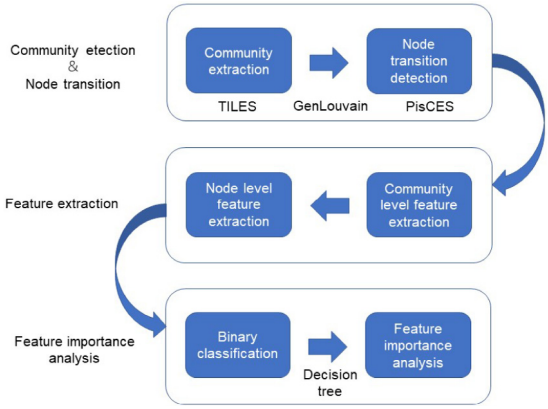


Fig. 2. Framework schematic.

Some researchers have realized that the node behavior is the driving force of community evolution, and community evolution plays a key role in temporal community detection. Therefore, some previous work has begun to use node structural features to enhance dynamic community detection [33].

In addition, there are some researchers trying to use node embedding model dynamic network [35]. They generally treat node embedding vectors as node structural features and then use the node embedding vectors to enhance dynamic community detection. However, this makes the node features unexplainable, as a result, we cannot understand how node structural features affect community-level evolution. Meanwhile, Yin et al. [36] used node average neighbor degree to enhance link prediction, but they did not discuss the general influence of node average neighbor degree on real-world data sets.

3. Proposed framework

In this section, we introduce how to find the most critical structural features that affect the transition behavior of nodes across the snapshots.

The proposed framework is depicted in Fig. 2, first of all, we use some temporal community detection methods to detect node community membership and node transition behaviors. Then, we extract the structural features of nodes and use them as classification features. We assume that the node community transition is only related to node structural features in the current snapshot, so we split snapshots in a network into adjacent snapshot pairs, and besides, different snapshot pairs in the same network are independent. We then treat the node community transition as a binary classification problem. For example, if a node i transfers its community membership between consecutive snapshots, then it is labeled $L_i = 1$. Finally, the most critical node structure features that affect the node community transition are analyzed.

To show the consistency of our proposed framework and experimental results, we select three popular and successful approaches for dynamic community detection: (1) TILES [29], which belongs to the evolutionary clustering framework. It effectively uses the network structure at time t and the community structure at the previous moment to detect the community at time t, which better community detection performance and lower computational complexity; (2) GenLouvain [37], which is a fast algorithm of modularity optimization for time-dependent networks. It generalizes the determination of community structure via quality functions to dynamic networks and could discover

Table 1
Notations and definitions.

Symbol	Feature	Description	Definition
$f1$	Community node number	Number of nodes within the community l at time t.	n_l^t
$f2$	Community edge number	Number of edges within the community l at time t.	e_l^t
$f3$	Intra community edges	Ratio of the total number of edges between the nodes inside the community ($e_l^t(in)$) to the number of nodes in the community.	$\frac{e_l^t(in)}{n_l^t}$
$f4$	Inter community edges	Ratio of the total number of edges of nodes connected outside the community ($e_l^t(out)$) to the number of nodes in the community.	$\frac{e_l^t(out)}{n_l^t}$
$f5$	Community activity	Ratio of the total number of connections made in the previous snapshot by the nodes of the community (a_l^t) to the number of nodes in the community.	$\frac{a_l^t}{n_l^t}$
$f6$	Community Conductance	Ratio of the number of edges in the community to the sum of degrees of the nodes in the community.	$\frac{e_l^t}{d_l^t}$
$f7$	Node degree	Sum of links connected to node i at time t.	e_i^t
$f8$	Node average neighbor degree	Average degree of node i's neighbors, where $N(i)^t$ are the neighbors of node i at time t and e_j^t is the degree of node j which belongs to $N(i)^t$.	$\frac{1}{\lvert N(i)^t\rvert}\sum_{j\in N(i)^t} e_j^t$
$f9$	Node closeness centrality	Measuring a node i's average path length to other nodes in community, where $C_{l,-i}^t$ is a set of all nodes in community l except i at time t and $d(i,j)$ is the distance between node i and j.	$\sum_{j\in C_{l,-i}^t}\frac{C_l^t}{d(i,j)}$
$f10$	Node betweenness centrality	Measuring a node i's importance in its community connectivity, where σ_{jk} is the total number of shortest paths from node j to node k and $\sigma_{jk}(i)$ is the number of those paths that pass through i	$\sum_{j,k\in C_{l,-i}^t}\frac{\sigma_{jk}(i)}{\sigma_{jk}}$

some important dynamic patterns; (3) PisCES [38], which is a global community detection method based on discovering persistent communities by eigenvector smoothing and combining information across a series of snapshots. It is also data-driven and can reveal dense communities that persist, merge, and diverge over time.

3.1. Notations and definitions

We use $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to denote a collection of the dynamic network, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of edges of the dynamic network. We fix the number of nodes in the network. Therefore, the nodes of the dynamic network do not change over time, which means that the size of the nodes collection $|\mathcal{V}| = N$ is a constant in the dynamic network. We use the changes in edges to represent the changes in the network. Therefore, a new node joins in the network can be treated as an isolated node with links to other nodes. Thus, the size of edges changes over time. We use $|\mathcal{E}|$ to represent the size of edges in the whole network and $|\mathcal{E}^t|$ to represent the size of edges in the tth snapshot. Furthermore, we also use $\mathcal{G} = (\mathcal{G}^1, \mathcal{G}^2, \ldots, \mathcal{G}^T)$ to represent a dynamic network, where T representing the number of network snapshots and $\mathcal{G}^t$ representing the tth snapshot in the network. We then use $\mathcal{C}^t = (c_1^t, \ldots, c_N^t)$ to represent the nodes community membership at snapshot t. More specifically, $c_i^t = k$ represents node i belong to community k at snapshot t.

3.2. Community detection and node transition detection

For the first step in our framework, we need to detect community structures and node transition behaviors from network triplet data. We select the following three methods, namely, TILES [29], GenLouvain [37] and PisCES [38].

- TILES [29] can solve both problems at the same time. TILES is a state-of-the-art evolutionary community detection algorithm. It proceeds to analyze an interactive stream: when a new link is generated, TILES uses a label propagation procedure to diffuse the changes to the node surroundings and adjust the neighborhood community membership. A node in TILES can belong to a community with two levels, i.e. peripheral level and core level. A node is a core node if it involves at least a triangle with other nodes in the same community, and it is a peripheral node if it is a one-hop neighbor of the core node. Only core nodes are allowed to spread community membership to their neighbors.
- GenLouvain [37] proposes a multislice generalization of modularity inspired by the equivalence between the modularity quality function (with a resolution parameter) and stability of communities under Laplacian dynamics. It can be used for multiple scales, time-dependent and multiplex networks. This metric can be effectively learned by any modularity optimization method.
- PisCES [38] extends the spectral clustering for dynamic networks through eigenvector smoothing, then it proposes an objective function based on the series of eigenvectors across the snapshots, and finally, an iterative algorithm is proposed for detecting the temporal communities.

It should be noted that TILES generates overlapping communities. It believes that overlapping communities represent different spheres of the social world of an individual. This brings us some troubles, because if a node belongs to different communities at the same time slice, how can we determine that this node changes its community membership in the next time slice? We believe that a node will transfer its community membership when it joins a new community, because a new community can represent a new social hub or new interest of an individual. Feature selection has always been an active and widely accepted method for enhancing the quality of data in machine learning and data mining.

3.3. Feature selection and extraction

Considering that the transfer behavior of a node is affected by the community in which it belongs to, we also introduce several community-level features. For example, if an online social group is not active enough, the members of the group will be more likely to join other groups. So we need to take community-level features into consideration, to verify its influence on the node community transition for the empirical evidence. Thus, our

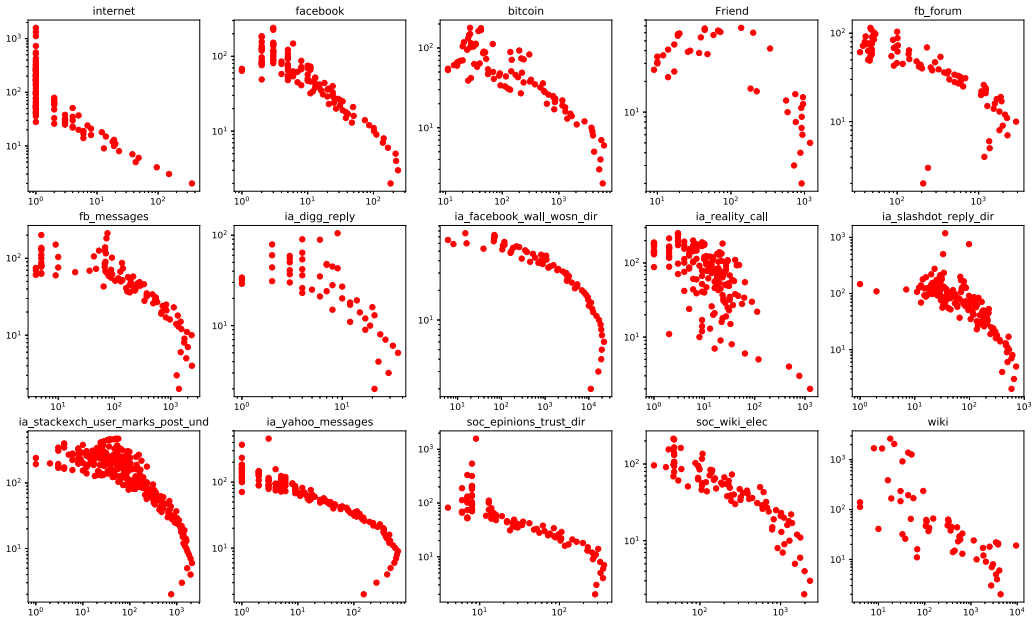


Fig. 3. The degree distributions of 15 dynamic networks.

Algorithm 1 Feature extraction

Input: A sequence of undirected graphs $\mathcal{G} = \mathcal{G}^1, \cdots, \mathcal{G}^T$ and the community assignment $\mathcal{C} = \mathcal{C}^1, \cdots, \mathcal{C}^T$

Output: Nodes feature set F and nodes label set L

for every graph $\mathcal{G}^t$ where $t \neq T$ **do**
  for every community $\mathcal{C}_l^t$ in $\mathcal{C}^t$ **do**
    Calculate community level features F_c
    for every node i in community $\mathcal{C}_l^t$ **do**
      Calculate node level features F_n
      Compose node i's feature sequence $F = F_c + F_n$
      if node i changes its community in $\mathcal{G}^{t+1}$ **then**
        node i's label L_i = 1
      else
        node i's label L_i = 0
      end if
    end for
  end for
end for

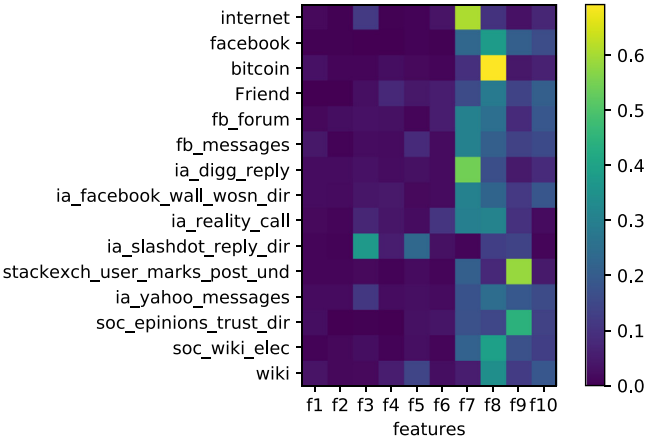


Fig. 5. The importance of different node characteristics or features on the dynamic behavior of 15 real dynamic networks based on the GenLouvain [37] method.

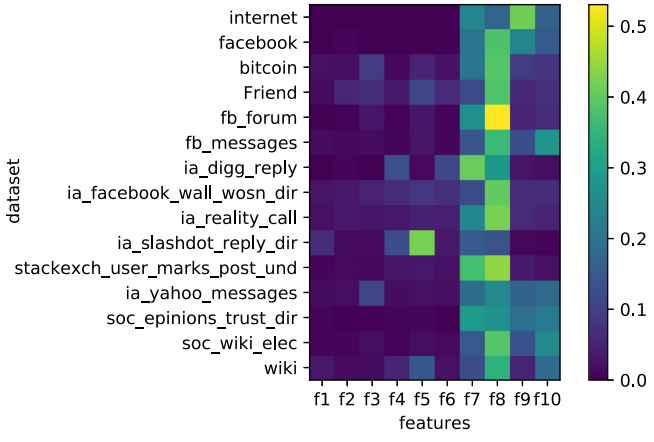


Fig. 4. The importance of different node characteristics or features on the dynamic behavior of 15 real dynamic networks based on the TILES [29] method.

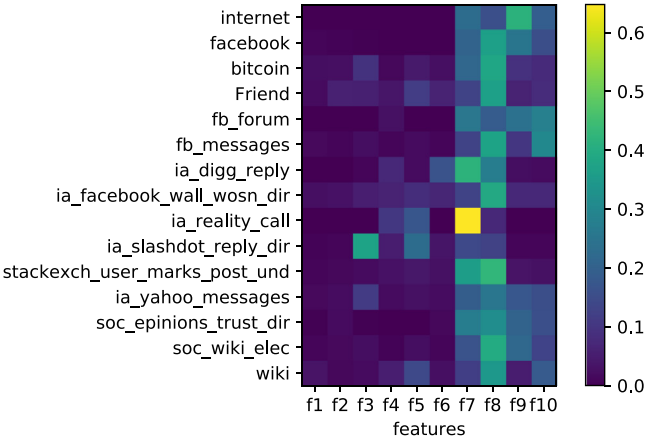


Fig. 6. The importance of different node characteristics or features on the dynamic behavior of 15 real dynamic networks based on the PisCES [38] method.

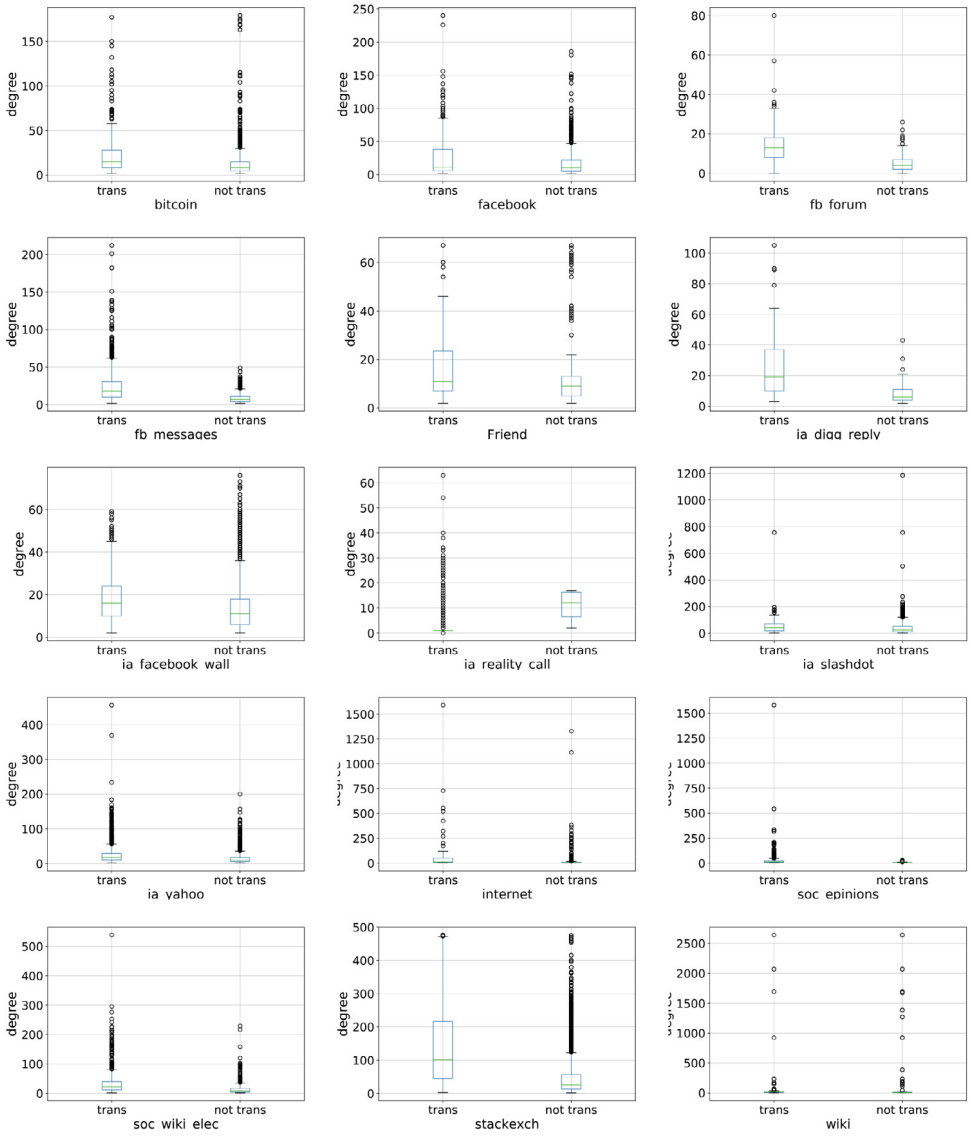


Fig. 7. Bitmaps of nodes degree on 15 data sets.

node features contain community structural features and node structural features in the network.

We use five community-level features to represent the state of the community in which the node is located, including the number of community nodes, the number of community edges, intra-community edges, inter-community edges and community activities. These features can be considered as higher-level local structural features of nodes. In addition, we use five node features to capture the node lower-level local structural features, including node degree, node average neighbor degree, node closeness centrality and node betweenness centrality. The detailed description of node features is listed in Table 1.

The feature extraction algorithm is provided in Algorithm 1. To calculate the node's features effectively, we calculate community-level features before calculating node features that are part of this community. And after splitting snapshots of a network into snapshot pairs, our feature extraction algorithm can run in parallel. This can make our algorithm suitable for a large scale network.

3.4. Feature importance analysis

We use decision tree [39] to solve the binary classification problem. Because a decision tree is a white-box algorithm, it can tell us which feature plays a more important role in the classification mission. Different from artificial neural networks and other black-box algorithms, it can output any information needed during the classification process.

A decision tree is a solution support tool that uses a tree-like graph or model of decisions. Each node of the decision tree is a "test" of the feature (e.g. whether a coin flip appears at the head or tail), each branch is the result of the test (e.g. the head and the tail of a coin flip are two branches), and each leaf node is a class label. The path from the root to the leaf represents the classification rule. Decision tree is an efficient algorithm for classification mission. The cost of using a tree is logarithmic, which makes it very fast in large data sets.

We use Gini importance or Mean Decrease in Impurity (MDI) to calculate each feature importance as the total decrease in node impurity [45]. Node impurity like Gini impurity is a computationally efficient approximation to the entropy, which can measure how well a potential split is separating the samples in a decision tree node. Decrease Impurity is defined as below:

$$\Delta i(s, r) = i(r) - p_L i(r_L) - p_R i(r_R), \tag{1}$$

where $i(r)$ is some impurity measure like Gini index, r represents a decision tree node, and r_L and r_R are the children of r. Besides,

Table 2
Description of data sets.

Name	Description	$\|\mathcal{V}\|$	$\|\mathcal{E}\|$
Internet	Internet [40] topology during 04/01/2004–04/04/2005.	33 936	104 824
Facebook	Facebook New Orleans networks [41] friends links during 06/08/2008–21/01/2009.	62 306	905 565
bitcoin	Who-trusts-whom network of people who trade using Bitcoin on Bitcoin OTC [42] during 09/11/2010–19/01/2016.	5881	35 592
Friend	Call logs of members of a young-family residential living community adjacent to a major research university in North America [43] during 10/07/2010–16/07/2011.	130	60 518
fb-forum	The Facebook-like Forum Network [44] during 15/05/2004–24/10/2004.	899	33 720
fb-messages	The Facebook-like Social Network [44] from an online community for students at University of California during 24/03/2004–22/10/2004.	1897	61 734
ia-digg-reply	A reply network of the social news website Digg [44] during 29/10/2008–13/11/2008.	30 397	87 627
ia-facebook-wall-wosn-dir	The Facebook friendship graph [44] during 15/05/2004–24/10/2004.	44 668	876 993
ia-reality-call	The MIT Reality mining a small set of human call logs data [44] during 24/09/2004–07/01/2005.	6810	52 050
ia-slashdot-reply-dir	Reply network of technology website Slashdot [44] during 01/12/2005–31/08/2006.	51 097	140 778
ia-stackexch-user-marks-post	User answering question network of Stack Overflow [44] during 03/10/2008–25/11/2011.	545 196	1 302 439
ia-yahoo-messages	The message network in Yahoo [44] with time presented by link sequences.	99 303	3 179 718
soc-epinions-trust-dir	Epinion who-trusts-whom network [44] with time presented by link sequences.	131 828	841 373
soc-wiki-elec	Wikipedia adminship election data [44] during 14/09/2004–05/01/2008.	8271	107 071
wiki	The Wikipedia links data [40] during 20/02/2001–06/12/2002.	329 623	39 953 145

$p_L = N_{r_L}/N_r$ and $p_R = N_{r_R}/N_r$, where N_r is the number of samples go through node r.

The normalized $\Delta i(s, r)$ for each feature can give us a kind of importance measure, and it is very computationally efficient for large scale data sets.

4. Experiment

In this section, we first introduce the details of the 15 real-world data sets used throughout this paper. Then we show the binary classification results in 15 data sets and our findings in feature importance experiment, that is, node degree and node average neighbor degree are the two most important structural features for node community transition.

4.1. Real-world data sets

The data sets used in the experiment contain different types of networks, for example, *social networks* with social account users as nodes and relationship as links, *friends cell phone call records networks* with phone owners as nodes and phone contacts as links, *who-trust-whom networks* with people as nodes and trust relationships as links and *tech-website answering questions networks* with website account as nodes and answering questions as links. The characteristics of the data sets and their sources are given in Table 2. Fig. 3 shows all of the node distributions of our data sets follow a power law.

4.2. Feature importance

We use the decision tree to process the node binary classification mission on 15 real-world data sets. As mentioned before, decision tree is an efficient algorithm for classification, so it is a good choice to execute binary classification in large data sets. Moreover, it allows us to calculate the feature importance in classification mission.

Figs. 4–6 show the importance of different node characteristics or features on the dynamic behavior of 15 real dynamic networks based on the TILES [29], GenLouvain [37] and PisCES [38], respectively. The results all show that node degree (f7) and node average neighbor degree (f8) are the two generally important features for node community transition on dynamic networks, which means that node degree and node average neighbor degree affect almost all the real-world data sets in community transition. Furthermore, node community activity plays an important role in slashdot reply data, but this feature is not applicable to all other data sets, which means that it does not have a general impact on community transition. Other community-level features, such as the number of community node, have no obvious guarantee against community transition. On the contrary, node-level features, such as node closeness centrality and node betweenness centrality, have a few intense on some data sets like internet, facebook and facebook messages. However, they did not work well on digg reply data or slashdot reply data. It also proves that the node closeness centrality and node betweenness centrality have no general influence on community transition.

Fig. 7 shows the bitmaps of node degree in different labels (1 represents nodes who transferred their community, 0 otherwise) on 15 data sets. As we can see, it shows an obvious pattern, that is, all nodes that transfer their communities have a higher degree than nodes that do not transfer their communities. This pattern proves that degree-corrected models like [7,46,47] have to be in conformity with the facts. Other features like node community activity, node closeness centrality and node betweenness centrality may play important roles in community transition in some data sets, but not all of them. Just as shown in Fig. 8,

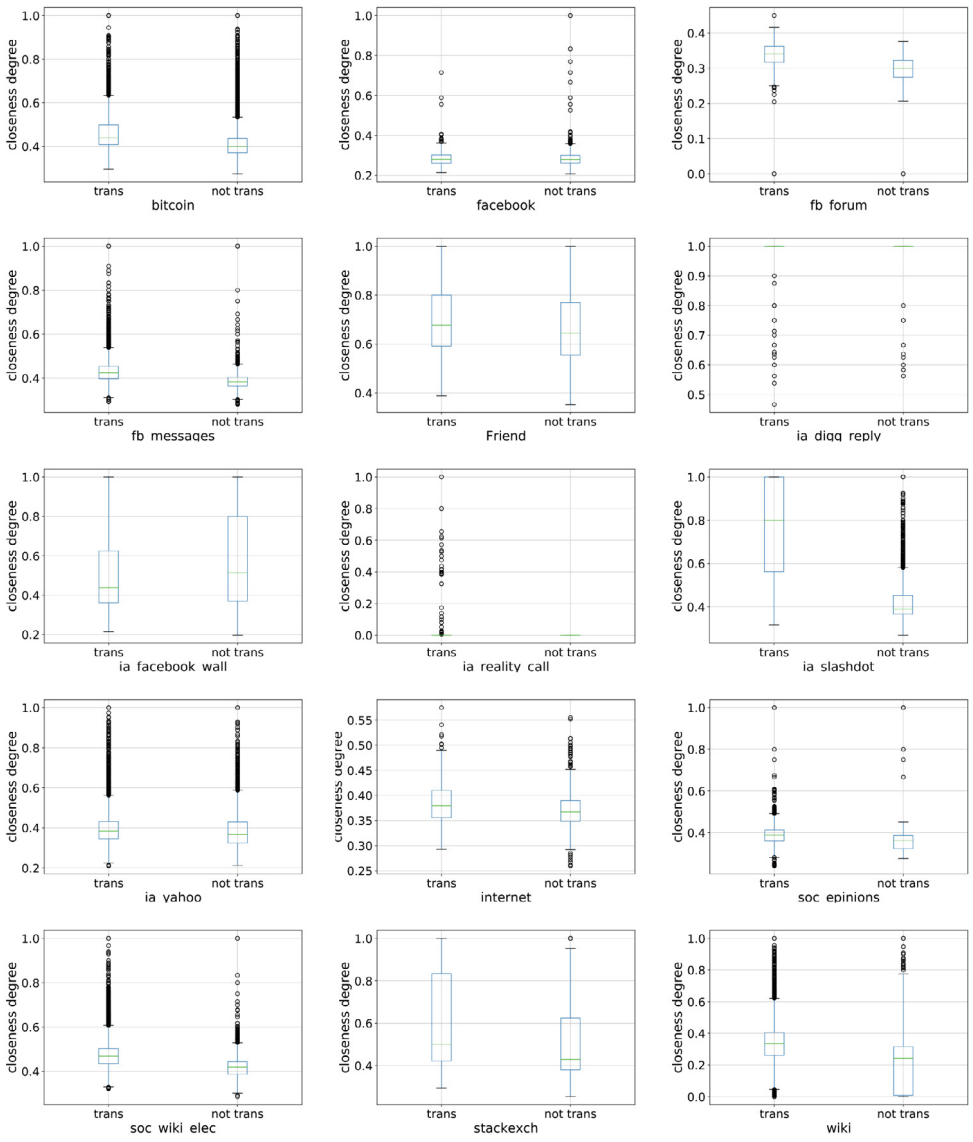


Fig. 8. Bitmaps of nodes closeness centrality on 15 data sets.

node closeness centrality does not show obvious pattern on all 15 data sets. Even in internet data, which shows the important instance of node closeness centrality, node closeness centrality still does not show significant differences between different labels. Furthermore, we find that the node average neighbor degree also plays an important role, or in other words, it plays a more important role than the node degree in almost all data sets we used. However, just as shown in Fig. 9, although the different labels on each data set are different, the bitmaps do not show an obvious consistent pattern. Through a case study on how the average neighbor degree of nodes affects the migration of a node community, it is found that it is still useful for dynamic community detection or community evolution.

Just as shown in Fig. 10(a), we chose three most representative data sets, namely, stackoverflow (left two columns), Friend (middle two columns) and facebook-wall-wosn (right two columns). The left column in every data set is the degree of nodes to which the nodes of communities have changed. The nodes that transferred their community have a larger degree than the nodes that did not. And as shown in Fig. 10(b), we also chose three most representative data sets, namely, facebook (left two columns), Friend (middle two columns) and facebook-wall-wosn (right two columns). The left column in every data set is the average neighbor degree of nodes that changed communities, which proves that nodes that transferred their communities have a larger average neighbor degree than the nodes that did not.

It is undoubtedly that the node degree can affect node community transition. Obviously, the higher the degree of nodes, the more likely it is to encounter nodes in other communities. And if the average neighbor degree of a node is larger, then the node is more likely to be affected by its neighbors. We will show a real-world case in the next section.

5. Case study

In this section, we use part of the DBLP data [15] to show the impact of node degree and node average neighbor degree on node community transition. DBLP is a well-studied data set in many research area, especially in complex network analysis. Our data is extracted from DBLP, and it contains the co-authorship information among the papers from 28 conferences over 10 years (1997–2007). These conferences cover three main research areas, including data mining, database and machine learning. Moreover, this data set has a ground truth, so we can extract the community membership of nodes without pre-processing this data.

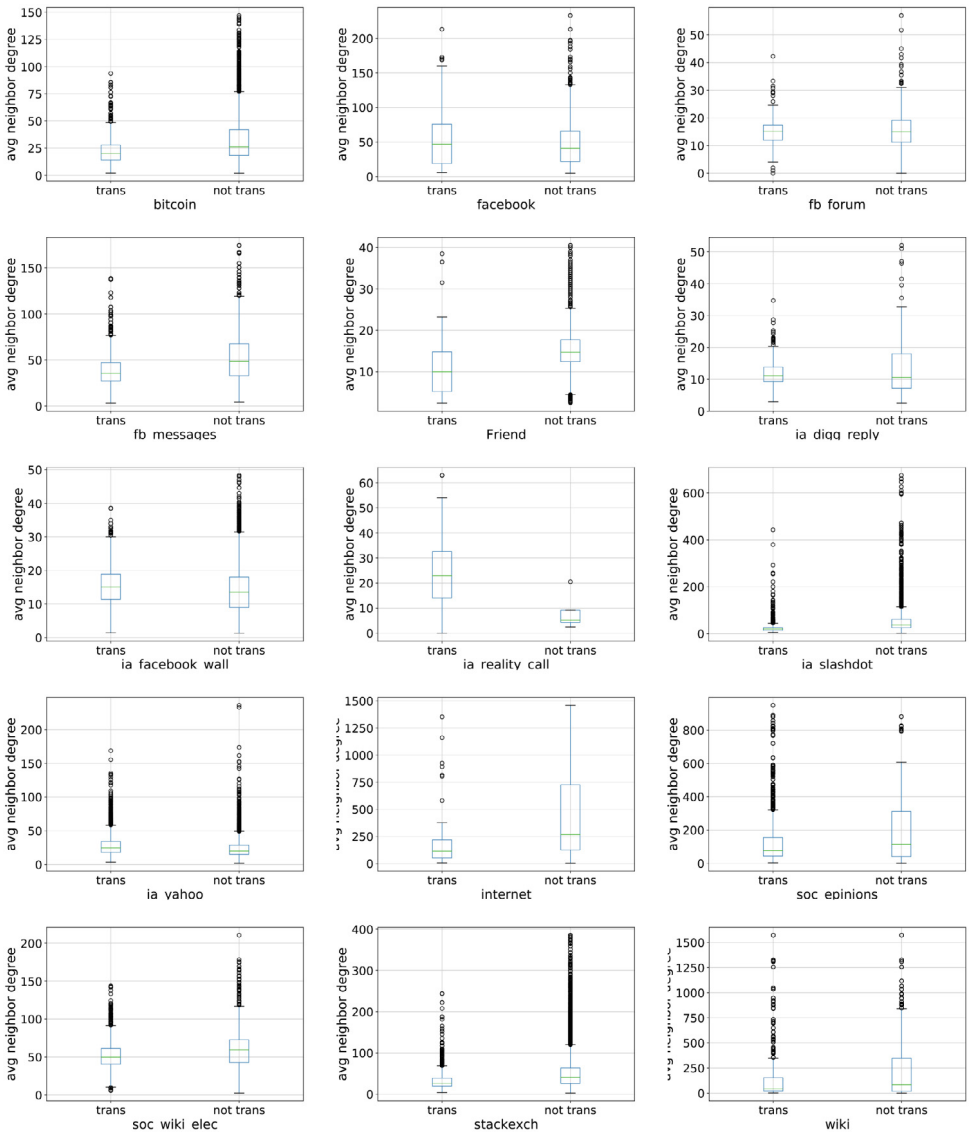
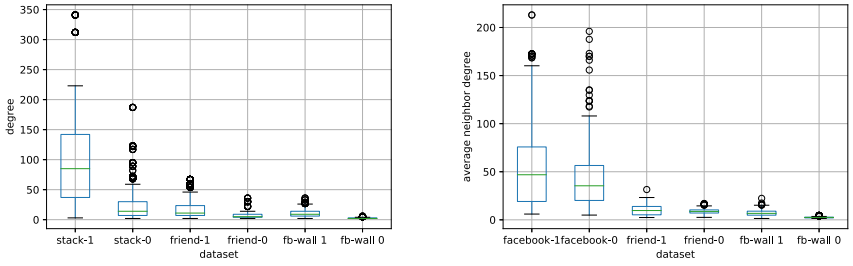


Fig. 9. Bitmaps of nodes average neighbor degree on 15 data sets. 1 in x-axis represents the bitmap of nodes transferring their communities, and 0 otherwise.



(a) Node degree (b) Node average neighbor degree

Fig. 10. Quartile map of node degree and node average neighbor degree in three data sets.

Fig. 11 shows the sample of DBLP data in year 2006–2007. The text of a node means 'average neighbor degree-author name', e.g. '4.82- *Shuicheng Yang*' means this node represents an author *Shuicheng Yang*, and its average neighbor degree is 4.82. And the scale of node represents node degree, i.e. a big node has more friends than a small node. The top two pictures show the influence of average neighbor degree. *Jun Yan*, *Zheng Chen* and *Ning Liu* are working on the database at previous snapshot (top-left), and they all have large average neighbor degree 9.33. They have a big degree friend *Shuicheng Yang* who is working on machine learning. Influenced by *Shuicheng Yang*, in the next snapshot (top-right) *Jun Yan*, *Zheng Chen* and *Ning Liu* change their research interest to machine learning, i.e. they published a paper about machine learning together in 2006. After investigation, we find that the above four authors jointly published an article on TKDE in

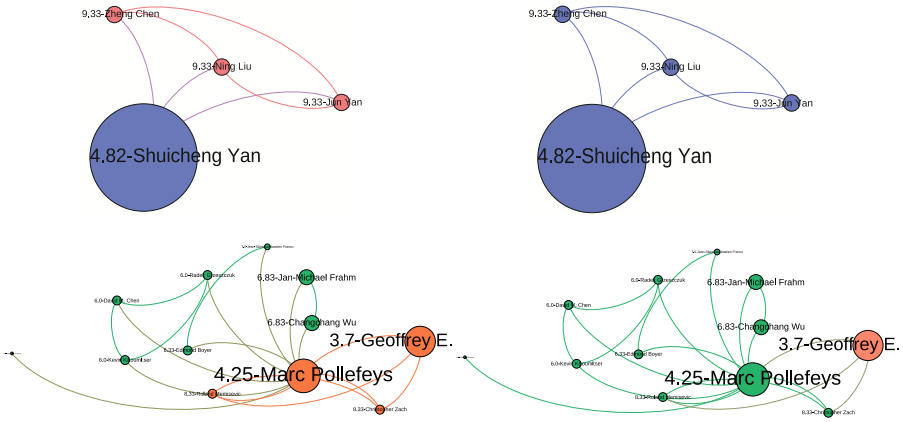


Fig. 11. Cases in DBLP shows the influence of node degree and average neighbor degree to node's community transition.

2006.[1] And the two pictures at the bottom show the influence of node degree. *Marc Pollefey* has a large degree, which means that he has many friends with this network. Most of his friends are working on data mining (bottom-left). Influenced by his friends, he changed his research area to data mining (bottom-right) in the next snapshot. More specifically, *Marc Pollefey* and one of his friends *Jan-Michael Frahm* published a paper together on EDGE in 2006.[2] Through these samples, we can intuitively understand the influence of the node degree and the node average neighbor degree on the transition of node community. However, more research on the node average neighbor degree is still needed to explore its impact mechanism on node community transition.

6. Conclusion

In this paper, we first consider the node's community transition as a binary classification problem. Through the analysis of 15 real-world dynamic networks, it is found that the degree and average neighbor degree of nodes are the two significant features that affect the pattern of node's community transition. In fact, we observe that node average neighbor degree is more important than the node degree, which is inconsistent with our previous understanding and also corrects our previous cognition of node transition factors. It has important reference meaning for the generation of dynamic networks and the detection of community structure. We also conduct a case study to explain the insurance against the node degree and node average neighbor degree to node community transition.

Unfortunately, the influence mechanism of the node average neighbor degree on the node community transition has not been found. This is the next step of our work. At the same time, one of the main directions of our future research is how to integrate our results with dynamic community detection methods, which is also what we will do next.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[1] Yan, Jun, et al. "Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing". IEEE transactions on Knowledge and Data Engineering 18.3 (2006): 320–333.

[2] Sinha, Sudipta N., et al. "GPU-based video feature tracking and matching". EDGE, workshop on edge computing using new commodity architectures. Vol. 278. 2006.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61902278, 51438009) and the National Key R&D Program of China (2018YFC0831000).

References

[1] M.E. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 167–256.
[2] N. Dakiche, F. Benbouzid-Si Tayeb, Y. Slimani, K. Benatchba, Tracking community evolution in social networks: A survey, Inf. Process. Manage. 56 (3) (2018) 1084–1102, http://dx.doi.org/10.1016/j.ipm.2018.03.005.
[3] M. Girvan, M.E. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. 99 (12) (2002) 7821–7826.
[4] K. Guruharsha, J.-F. Rual, B. Zhai, J. Mintseris, P. Vaidya, N. Vaidya, C. Beekman, C. Wong, D.Y. Rhee, O. Cenaj, et al., A protein complex network of Drosophila melanogaster, Cell 147 (3) (2011) 690–703.
[5] G.A. Pagani, M. Aiello, The power grid as a complex network: a survey, Physica A 392 (11) (2013) 2688–2700.
[6] D. Džamić, D. Aloise, N. Mladenović, Ascent–descent variable neighborhood decomposition search for community detection by modularity maximization, Ann. Oper. Res. 272 (1–2) (2019) 273–287.
[7] B. Karrer, M.E. Newman, Stochastic blockmodels and community structure in networks, Phys. Rev. E 83 (1) (2011) 016107.
[8] Y.-M. Wen, L. Huang, C.-D. Wang, K.-Y. Lin, Direction recovery in undirected social networks based on community structure and popularity, Inform. Sci. 473 (2019) 31–43.
[9] D. He, Z. Feng, D. Jin, X. Wang, W. Zhang, Joint identification of network communities and semantics via integrative modeling of network topologies and node contents, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 116–124.
[10] E.M. Airoldi, D.M. Blei, S.E. Fienberg, E.P. Xing, Mixed membership stochastic blockmodels, J. Mach. Learn. Res. 9 (Sep) (2008) 1981–2014.
[11] M. Qiao, J. Yu, W. Bian, Q. Li, D. Tao, Improving stochastic block models by incorporating power-law degree characteristic, in: IJCAI, 2017, pp. 2620–2626.
[12] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (3–5) (2010) 75–174.
[13] S. Fortunato, D. Hric, Community detection in networks: A user guide, Phys. Rep. 659 (2016) 1–44.
[14] H. Liao, M.S. Mariani, M. Medo, Y.-C. Zhang, M.-Y. Zhou, Ranking in evolving complex networks, Phys. Rep. 689 (2017) 1–54.
[15] T. Yang, Y. Chi, S. Zhu, Y. Gong, R. Jin, Detecting communities and their evolutions in dynamic social networks—a Bayesian approach, Mach. Learn. 82 (2) (2011) 157–189.
[16] G. Palla, A.-L. Barabási, T. Vicsek, Quantifying social group evolution, Nature 446 (7136) (2007) 664.
[17] Y. Sun, J. Tang, L. Pan, J. Li, Matrix based community evolution events detection in online social networks, in: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom, SmartCity, IEEE, 2015, pp. 465–470.
[18] M.-S. Kim, J. Han, A particle-and-density based evolutionary clustering method for dynamic networks, Proc. VLDB Endow. 2 (1) (2009) 622–633.
[19] D. Chakrabarti, R. Kumar, A. Tomkins, Evolutionary clustering, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2006, pp. 554–560.

[20] X. Fan, L. Cao, R.Y. Da Xu, Dynamic infinite mixed-membership stochastic blockmodel, IEEE Trans. Neural Netw. Learn. Syst. 26 (9) (2014) 2072–2085.
[21] X. Tang, C.C. Yang, Detecting social media hidden communities using dynamic stochastic blockmodel with temporal dirichlet process, ACM Trans. Intell. Syst. Technol. (TIST) 5 (2) (2014) 36.
[22] S. Sengupta, Y. Chen, A block model for node popularity in networks with community structure, J. R. Stat. Soc. Ser. B Stat. Methodol. 80 (2) (2018) 365–386.
[23] L. Yu, W.H. Woodall, K.-L. Tsui, Detecting node propensity changes in the dynamic degree corrected stochastic block model, Social Networks 54 (2018) 209–227.
[24] L. Huang, C.-D. Wang, H.-Y. Chao, A harmonic motif modularity approach for multi-layer network community detection, in: 2018 IEEE International Conference on Data Mining, ICDM, IEEE, 2018, pp. 1043–1048.
[25] J. Liu, T. Liu, Detecting community structure in complex networks using simulated annealing with k-means algorithms, Physica A 389 (11) (2010) 2300–2309.
[26] U. Von Luxburg, A tutorial on spectral clustering, Stat. Comput. 17 (4) (2007) 395–416.
[27] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, P. Zhang, Spectral redemption in clustering sparse networks, Proc. Natl. Acad. Sci. 110 (52) (2013) 20935–20940.
[28] E.G. Tajeuna, M. Bouguessa, S. Wang, Tracking communities over time in dynamic social network, in: International Conference on Machine Learning and Data Mining in Pattern Recognition, Springer, 2016, pp. 341–345.
[29] G. Rossetti, L. Pappalardo, D. Pedreschi, F. Giannotti, Tiles: an online algorithm for community discovery in dynamic social networks, Mach. Learn. 106 (8) (2017) 1213–1241.
[30] L. Yang, Y. Guo, D. Jin, H. Fu, X. Cao, 3-in-1 correlated embedding via adaptive exploration of the structure and semantic subspaces, 2017, pp. 3613–3619.
[31] S. Yang, H. Koeppl, A Poisson gamma probabilistic model for latent node-group memberships in dynamic networks, 2018, pp. 4366–4373, arXiv: 1805.11054.
[32] P. Bródka, S. Saganowski, P. Kazienko, GED: the method for group evolution discovery in social networks, Soc. Netw. Anal. Min. 3 (1) (2013) 1–14.
[33] D. Greene, D. Doyle, P. Cunningham, Tracking the evolution of communities in dynamic social networks, in: 2010 International Conference on Advances in Social Networks Analysis and Mining, IEEE, 2010, pp. 176–183.
[34] S. Asur, S. Parthasarathy, D. Ucar, An event-based framework for characterizing the evolutionary behavior of interaction graphs, ACM Trans. Knowl. Discov. Data 3 (4) (2009) 16.
[35] L. Zhou, Y. Yang, X. Ren, F. Wu, Y. Zhuang, Dynamic network embedding by modeling triadic closure process, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[36] H. Yin, A.R. Benson, J. Leskovec, The local closure coefficient: a new perspective on network clustering, in: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, ACM, 2019, pp. 303–311.
[37] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, Science 328 (5980) (2010) 876–878.
[38] F. Liu, D. Choi, L. Xie, K. Roeder, Global spectral clustering in dynamic networks, Proc. Natl. Acad. Sci. 115 (5) (2018) 927–932.
[39] M. Dumont, R. Marée, L. Wehenkel, P. Geurts, Fast multi-class image annotation with random subwindows and multiple output randomized trees, in: Proc. International Conference on Computer Vision Theory and Applications, Vol. 2, VISAPP, 2009, pp. 196–203.
[40] A. Mislove, Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems (Ph.D. thesis), Rice University, Department of Computer Science, 2009.
[41] B. Viswanath, A. Mislove, M. Cha, K.P. Gummadi, On the evolution of user interaction in facebook, in: Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks, WOSN'09, 2009, pp. 37–42.
[42] S. Kumar, F. Spezzano, V. Subrahmanian, C. Faloutsos, Edge weight prediction in weighted signed networks, in: Data Mining (ICDM), 2016 IEEE 16th International Conference on, IEEE, 2016, pp. 221–230.
[43] N. Aharony, W. Pan, C. Ip, I. Khayal, A. Pentland, Social fMRI: Investigating and shaping social mechanisms in the real world, Pervasive Mob. Comput. 7 (6) (2011) 643–659.
[44] R.A. Rossi, N.K. Ahmed, The network data repository with interactive graph analytics and visualization, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015, pp. 4292–4293.
[45] B.H. Menze, B.M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F.A. Hamprecht, A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data, BMC Bioinform. 10 (1) (2009) 213.
[46] J.D. Wilson, N.T. Stevens, W.H. Woodall, Modeling and detecting change in temporal networks via a dynamic degree corrected stochastic block model, 2016, arXiv preprint arXiv:1605.04049.
[47] D. Jin, Z. Chen, D. He, W. Zhang, Modeling with node degree preservation can accurately find communities, in: Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015, pp. 160–167.

Tianpeng Li received the B.S. degree from school of software engineering, Tianjin university in 2017. He is currently pursuing the M.S. degree with the School of College of Intelligence and Computing, Tianjin University. His research interests include machine learning, dynamic complex network analysis, dynamic community detection and community evolution in dynamic network and probabilistic graphical models and its applications in computer science.

Wenjun Wang is currently a Professor at the School of College of Intelligence and Computing, Tianjin University, Chief expert of major projects of the National Social Science Foundation, the big data specially-invited expert of Tianjin Public Security Bureau and the director of the Tianjin Engineering Research Center of Big Data on Public Security. His research interests include computational social science, large-scale data mining, intelligence analysis and multi-layer complex network modeling. He was the principal investigator or was responsible for more than 50 research projects, including the Major Project of National Social Science Fund, the Major Research Plan of the National Natural Science Foundation, the National Science2013technology Support Plan Project of China, etc. He has published more than 50 papers on main international journals and conferences.

Xunxun Wu received the B.S. degree in mathematics from Shandong University, Jinan, China, in 2016. She is currently pursuing the M.S. degree with the School of College of Intelligence and Computing, Tianjin University, Tianjin, China. Her current research interests include complex network analysis and data mining, and currently working on community detection, community evolution in dynamic networks, and probabilistic graphical model.

Huaming Wu received the B.E. and M.S. degrees from Harbin Institute of Technology, China in 2009 and 2011, respectively, both in electrical engineering. He received the Ph.D. degree with the highest honor in computer science at Free University of Berlin, Germany in 2015. He is currently an associate professor in the Center for Applied Mathematics, Tianjin University. His research interests include mobile cloud computing, edge computing, fog computing, internet of things (IoTs), and deep learning.

Pengfei Jiao received the Ph.D. degrees in computer science from Tianjin University, Tianjin, China, in 2018. He is a lecture with the Center of Biosafety Research and Strategy of Tianjin University. His current research interests include complex network analysis and data mining, and currently working on community detection and link predication, community evolution in dynamic networks, network embedding and applications of statistical network model.

Yandong Yu, an associate professor, works in the Department of Computer Science of Jining Normal University, Wulanchabu, Inner Mongolia. She is currently a visiting scholar in the School of College of Intelligence and Computing at tianjin university. In 2003, she obtained the bachelor degree of computer science and technology from Tianjin Normal University. In 2011, she obtained the master degree of computer technology engineering from Inner Mongolia University. Her research interests include big data analysis, complex networks, network security and so on.