Notes on DCDID:

- Identifying communities in dynamic networks is essential for exploring the latent network structures, understanding network functions, predicting network evolution, and discovering abnormal network events

- In this paper, we present a dynamic community detection framework based on information dynamics and develop a dynamic community detection algorithm called DCDID (dynamic community detection based on information dynamics), which uses a batch processing technique to incrementally uncover communities in dynamic networks.

- DCDID employs the information dynamics model to simulate the exchange of information among nodes and aims to improve the efficiency of community detection by filtering out the unchanged sub-graph.

Theme:

- This model is based on the features to automatically display the community structure by simulating information exchange between members. The authors devised a novel technique based on information dynamics for gaining insights regarding community division in dynamic networks, with the core premise being to examine an accommodative dynamical system and investigate its information dynamics over time.

- In particular, in an interpersonal network, people with similar interests or features are more likely to interact with others, and the propagation of information between them tends to be more frequent. With the diffusion and interaction of information, people in the same community have almost the equivalent amount of information, whereas those in diverse communities have different amounts of information. Over time, the information dynamics on the network reaches the steady state, and the communities can be naturally uncovered by counting the amount of information of nodes in the network.

Example:

In this dynamic artificial network, we take a company as an example to present the process of dynamic community detection based on information dynamics,which comprises the following stages:

- First, everyone possesses his/her own knowledge as initialinformation because of distinct occupations (i.e.,v=0.66,u1=0.5), as shown in Figure 1a.
- Then, the information spreads through the topological structure of the network (seeEquation (4)).
- For example, user v shares its information with the connected neighbors u1−u4, as shown in Figure 1b. Over time, the amount of information exchanged between people tends to zero and the information dynamics reaches the steady state. Next, the communities are naturally revealed by

computing the different information in the network. Figure 1b shows the community structure detected by the information dynamics model in the time slice T0. Building upon the information of communities detected at the time sliceT0, an incremental community discovery framework is adopted for the subsequent snapshot networks (see Figure 2), which includes adding nodes, deleting nodes,adding edges, and deleting edges events.

- Figure 1c–f demonstrates that the addition and deletion of nodes and edges may lead to changes in the network structure.
- For example, the addition of useru5causes the two communities to merge into one community, as shown in Figure 1c.
- Similarly, Figure 1d, represent that the deletion of users u6 and u7 results in a split in the community.

Dynamic Community Detection Algorithm:
(1)Community Detection based on Information Dynamics: Based on the information dynamics models, we identify the community structure by simulating the interaction of information on the networks.
In the beginning, each node is provided initial information in light of the local topology features .

Definition 3. *(Information) Let $G_t = (V_t, E_t)$ be an undirected network at time step t, and the information of vertex u is defined as follows:*

$$I_u = \frac{D_u}{D_{max}} \tag{3}$$

where D_u represents the degree of vertex u, and D_{max} denotes the maximum degree of the network G_t.

Then, the information diffuses in the network and every node is constantly interacting with neighbor nodes. The exchange of information between nodes in the same community is more frequent than that in different communities. At each step, every node updates its information based on the information dynamics models (cf. Equation 7).

Information Propagation. Finally, by considering the information volume and the information loss described above, the information dynamics of a node v over time is provided by the following:

$$I_{v(t+1)} = I_{v(t)} + \sum\nolimits_{u \in N(v)} (I_{u \to v} - I_{(u \to v)\_cost}) \tag{7}$$

where $I_{v(t)}$ represents the information of node v at time step t, and the second term of this formula denotes the information that is acquired from its neighbors. As we can see, the information of node v at time step $t+1$ includes the information at time step t and the information obtained from its neighbors at time step $t+1$. With the evolution of time, the information propagated tends to zero. Eventually, the information in the network will reach equilibrium, and the communities can be uncovered naturally.

As time evolves, the exchange of information between nodes tends to zero, and the information dynamics of each node in the network reaches the convergent state. Finally, the amount of information for each node in the same community is basically the same, and the information on each node in different communities is different. Therefore, we can naturally uncover the communities by considering the amount of information for each node.

(2) Dynamic Community Detection: DCDID mainly consists of the three steps: initialcommunity structure detection, calculation of subgraphs that have changed, and incremental community identification.

(a) Initial Community Structure Detection. The initial community structure is the community partition of the network at time sliceT0. There is no prior information about community structure in the initial time slice, so it is necessary to perform community detection on the entire network. We use the community detection based on information dynamics (CDID) to identify the community structure of the initial network at time sliceT0. The CDID algorithmis given in the appendices

b)Changed Subgraphs. Considering the operations that may cause changes in the community structure, we divide the events that change the network into four categories: adding nodes,deleting nodes, adding edges, and deleting edges.

Algorithm 3–6 in Appendix A show thespecific process, and each type of event returns a subgraph that may change

(b) Incremental Community Identification. At present, most incremental dynamic community detection methods adopt the fine-grained processing method, which processes an event when an event is generated. For example, when a node is added to the network, the nodeis detected. The advantage of this design is that the processing of events is takes place in real time, but the disadvantage is that it increases the computational complexity. Here, we employ a batch-based incremental community detection method. Based on the obtained subgraphs that may change, we employ the information dynamics model to incrementally detect the communities.

Algo:
Input: Gt = (Vt, Et)
Output: Ct
1: //Initialization of information
2: **for** each node v 2 Vt **do**
3: **for** each node u 2 N(v) **do**
4: compute the JSvu, CSuv using Equation (1)–(2)

Definition 1. *(Jaccard similarity coefficient [33]) Let $G_t = (V_t, E_t)$ be an undirected network at time steps t. The Jaccard similarity coefficient of two nodes v and u is defined as follows:*

$$JS_{vu} = \frac{| \tau(v) \bigcap \tau(u) |}{| \tau(v) \bigcup \tau(u) |} \tag{1}$$

where $u \in V$, $v \in V$, $\tau(u) = N(u) \cup \{u\}$, and $N(u)$ is the set of adjacent nodes of node u .

In real life, social networks often include strong and weak relationships, which play a significant role in information propagation and community formation. To describe this relationship, we use contact strength to represent the degree of tightness between nodes in a given network. Here, we use triangles to formalize the definition of the contact strength because the triangle structure can better characterize the tightness of the vertices.

Definition 2. *(Contact strength) Let $G_t = (V_t, E_t)$ be an undirected network at time step t, and the contact strength of vertex v on vertex u is defined as follows:*

$$CS_{uv} = \frac{|N(u) \bigcap N(v)|}{T_u} \tag{2}$$

where T_u denotes the number of triangles for vertex u, and the intersection between $N(u)$ and $N(v)$ represents the amount of triangles shared by two nodes u and v.

Here, CS_{uv} is an asymmetric function, in other words, the values of CS_{uv} and CS_{vu} may not be equal. For example, everyone has their own circle of friends, and the contact strength between two people is likely to be unequal.

In the real world, the more friends a person has, the more resources he has, so he may obtain more information. To describe the information of nodes in the network, we use the degree of nodes to characterize the initial information of the nodes.

5: **end for**
6: compute the Iv using Equation (3)

Definition 3. *(Information) Let $G_t = (V_t, E_t)$ be an undirected network at time step t, and the information of vertex u is defined as follows:*

$$I_u = \frac{D_u}{D_{max}} \tag{3}$$

where D_u represents the degree of vertex u, and D_{max} denotes the maximum degree of the network G_t.

7: **end for**
8: //Information dynamic interaction.

Information Dynamic Model

To reveal the communities in the dynamic networks, we begin to construct the information dynamic model, which includes three parts: information propagation volume, information loss, and propagation model.

We assume that everyone may obtain knowledge from their neighbors and disseminate information to them based on studies on the patterns of information dissemination among individuals in the actual world. The local topology of a network, such as the degree of a node, similarities, and connection strengths of nodes, has a significant impact on information dispersion. Those, for example, prefer to speak with people with whom they share deep ties or shared interests. To represent the amount of information diffusion, we use node similarity, connection strength, and information difference to simulate the amount of information diffusion in a more realistic way.

Compute the flow of info
9: **while** true **do**
10: Imax = 0
11: **for** each node v 2 Vt **do**
12: **for** each node u 2 N(v) **do**
13: compute Iu!v using Equation (4)–(5)

diffusion. Formally, let $I_{u \to v}$ represent the information that a node v obtains from its neighbor u, which is defined as follows:

$$I_{u \to v} = f(I_u - I_v) JS_{uv} CS_{uv} \quad (4)$$

where JS_{uv} denotes the Jaccard similarity coefficient between node u and node v, and CS_{uv} represents the contact strength of node v on node u. The coupling function $f(\cdot)$ denotes the information that can be disseminated from the u to v, which is defined as follows:

$$f(I_u - I_v) = \begin{cases} e^{(I_u - I_v)} - 1 & I_u - I_v \geq 0 \\ 0 & I_u - I_v < 0. \end{cases} \quad (5)$$

We can see that the nodes with a large information volume are more likely to diffuse and affect the nodes with a small information volume. When the information of I_u and I_v are close to equal, the amount of information passed between them tends to zero.

14: compute I(u!v)_cost using Equation (6)
15: **end for**
16: compute Iv(t+1) using Equation (7)

Information Loss. In the real world, because of the influence of complex environments, loss may occur during information dissemination. As an example, if people are familiar or interested in the information diffused, they may understand and propagate it more easily. Conversely, we may ignore, misunderstand or even lose information. To reflect the loss of information in a more realistic and accurate way, we use the volume of information and the topological features for its characterization. Let $I_{(u\to v)\_cost}$ denote the loss of information, which is defined as follows:

$$I_{(u\to v)\_cost} = \frac{Avg_{s(v)}}{Avg_{d(v)}} f(I_u - I_v) * (1 - JS_{uv}) \tag{6}$$

where the first item of the formula characterizes the local topological feature, which consists of the local average similarity and local average degree. $I_{(u\to v)\_cost}$ is positively correlated with coupling function $f(\cdot)$ and negatively correlated with JS_{uv}. Therefore, the larger the information volume to be diffused, the greater the information loss is, and the more similar the communication objects are, the smaller the information loss is.

Information Propagation. Finally, by considering the information volume and the information loss described above, the information dynamics of a node v over time is provided by the following:

$$I_{v(t+1)} = I_{v(t)} + \sum\nolimits_{u\in N(v)} (I_{u\to v} - I_{(u\to v)\_cost}) \tag{7}$$

where $I_{v(t)}$ represents the information of node v at time step t, and the second term of this formula denotes the information that is acquired from its neighbors. As we can see, the information of node v at time step $t+1$ includes the information at time step t and the information obtained from its neighbors at time step $t+1$. With the evolution of time, the information propagated tends to zero. Eventually, the information in the network will reach equilibrium, and the communities can be uncovered naturally.

```
Iin = Iu!v    I(u!v)_cost
if Iin > Imax then
Imax = Iin
end if
end for
// the balanced state
if Imax < Threshold then
Break
end if
end while
// Find communities Ct
for each node v 2 Vt do
if v /2 Ct then
for each node u 2 N(v) do
if jIv    Iuj < Threshold then
u   > Cv
else
u   > Cu
end if
end for
end if
```