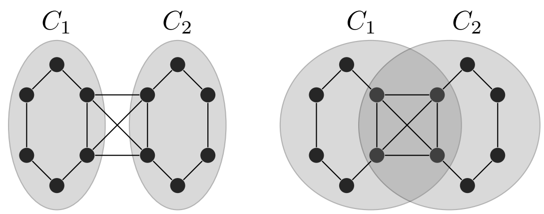
Online Resourse 2

immediate

March 1, 2019

1 Refinement of Communities

Overlap between Communities To illustrate the overlap function θ, consider the subgraphs C_1 and C_2 as depicted in the following pictures.



If we assume that all the edges have the same weight, then in the left picture, $\theta(C_1, C_2) = 4/12 = 1/3$. However, in the right one, $C_1 \cap C_2 \neq \varnothing$. So, a_{uv} will be added twice for $u, v \in C_1 \cap C_2$. Then we have $\theta(C_1, C_2) = 8/12 = 2/3$. This means θ increases as the size of $C_1 \cap C_2$ increases. However, this process is not as easier as it seems. We need to look at it more closely.

Refinement Procedure Assume that $labels$ contains all the labels of the initial communities stored in $\mathcal{K}$. Choose i randomly from $labels$. Let us denote the community $\mathcal{K}[i]$ by C_i. Let $nbrc\_labels$ be the set of all the labels of the neighboring communities of C_i. This means, for any $j \in nbrc\_labels$, $E(C_i \cap C_j) \neq \varnothing$. Now we find hoc_labels as a subset of $nbrc\_labels$ in such a way that for all $j \in hoc\_labels$, $\theta(C_i, C_j) > \theta_0$. In many cases, however, it may happen that $\theta(C_i, C_j) \leq \theta_0$, for all $j \in nbrc\_labels$. Additionally, a user might be interested in the communities higher than a particular size, say n_{min}. For such cases, we choose hoc_labels in such a way that for all $j \in hoc\_labels$, $\theta(C_i, C_j)$ is highest. Once the set hoc_labels is in hand we merge C_i into C_j for each $j \in hoc\_labels$ and update the memberships of each of the vertices of C_i accordingly. After this C_i is removed from $\mathcal{K}$ and i is removed from $labels$. This procedure is repeated as long as $labels$ is nonempty. The details procedure is provided in **Procedure** CommRefine.

Procedure CommRefine($\mathcal{K}, m, \theta_0, n_{min}$)

Result: $\mathcal{K}, m$

$labels \leftarrow 1 : |\mathcal{K}|$
repeat
    Let i be chosen from $labels$ randomly and let $nbrc\_labels$ be the labels of all those communities in $\mathcal{K}$ that have at least one neighbor in C_i.
    $max\_ov \leftarrow 0, hoc\_labels \leftarrow \varnothing$
    for $j \in nbrc\_labels$ **do**
        if $\theta(C_i, C_j) > \theta_0$ **then**
            $hoc\_labels \leftarrow hoc\_labels \cup j$
        end
    end
    if $hoc\_labels = \varnothing$ **then**
        Choose hoc_labels as the subset of $nbrc\_labels$ in such a way that for each $j \in hoc\_labels$, $\theta(C_i, C_j)$ is highest.
    end
    for $j \in hoc\_labels$ **do**
        for $v \in C_i$ **do**
            $C_j \leftarrow C_j \cup v$
            $m_v \leftarrow m_v - i$
            $m_v \leftarrow m_v \cup j$
        end
    end
    if $hoc\_labels \neq \varnothing$ **then**
        $\mathcal{K} \leftarrow \mathcal{K} - C_i$
    end
    $labels \leftarrow labels - i$
until $labels = \varnothing$