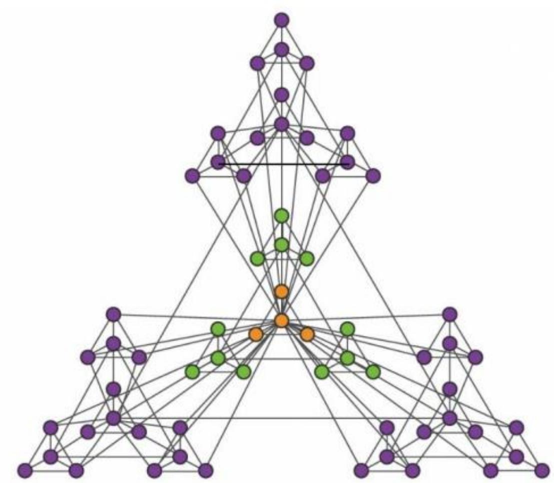
HOMEWORK 03: Frequent Graph Pattern Mining

Student name: Ngô Văn Trung Nguyên

StudentID: 20127054

<u>Problem 1. Hierarchical Networks:</u>

-Calculate the degree exponent of the hierarchical network shown in figure below.



- We can see that the largest hub at the center of the network acquires 3^n links during the n^{th} iteration.
- The hierarchical model generates a scale-free network with degree exponent:

$$\boldsymbol{\gamma = 1 + \frac{ln4}{ln3} = 2.262}$$

<u>Problem 2. Communities on a Circle</u>

Consider a one dimensional lattice with N nodes that form a circle, where each node connects to its two neighbors. Partition the line into nc consecutive clusters of size$N_c = {}^{N}/_{n_c}$.

a) Calculate the modularity of the obtained partition.

- We have: $M_c = \frac{L_c}{L} - (\frac{k_c}{2L})^2$ (where L_c is the total number of links within the community C_c and k_c is the total degree of the nodes in this community).
- For one community: $M_c = \frac{L_c}{L} - (\frac{k_c}{2L})^2 = \frac{N_c - 1}{N} - (\frac{2N_c}{2N})^2$
  - Number of links L_c in comunity is the number of nodes in that comunity – 1
  - All nodes have degree 2 so $k_c = number\ of\ nodes\ in\ that\ community * 2$
  - This is a circle, so L = N.

- Because $N_c = \frac{N}{n_c}$, we have:

$$M_c = \frac{N_c - 1}{N} - \left(\frac{2N_c}{2N}\right)^2 = \frac{\frac{N}{n_c} - 1}{N} - \left(\frac{\frac{2N}{n_c}}{2N}\right)^2 = \frac{N - n_c}{N * n_c} - \frac{1}{n_c}^2 = \frac{1}{n_c} - \frac{1}{N} - \frac{1}{{n_c}^2}$$

- Because all communities are the same, we have the partition's modularity:

$$\boldsymbol{M} = \sum_{c=1}^{n_c} \left[\frac{L_c}{L} - (\frac{k_c}{2L})^2\right] = \left[\frac{1}{n_c} - \frac{1}{N} - \frac{1}{{n_c}^2}\right] * \boldsymbol{n_c} = \boldsymbol{1} - \frac{\boldsymbol{n_c}}{\boldsymbol{N}} - \frac{\boldsymbol{1}}{\boldsymbol{n_c}}$$

b) According to the Maximum Modularity Hypothesis, the maximum of Mc corresponds to the best partition. Obtain the community size nc corresponding to the best partition.

- To have M reach maximum, we need to find the maximum of

$$M = 1 - \frac{n_c}{N} - \frac{1}{n_c}$$

- Diffentiate:

$$\frac{dM}{dn_c} = 0 - \frac{1}{N} + \frac{1}{{n_c}^2} = 0$$

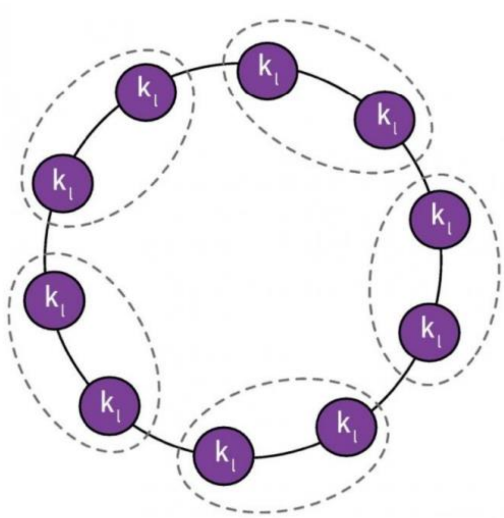
$$\rightarrow \frac{1}{{n_c}^2} = \frac{1}{N}$$

$$\rightarrow n_c = \sqrt{N}$$

- Therefore, $\boldsymbol{n_c} = \sqrt{\boldsymbol{N}}$.

Problem 3. Modularity Resolution Limit

Consider a network consisting of a ring of n_c cliques, each clique having N_c nodes and $\frac{m(m-1)}{2}$ links. The neighboring cliques are connected by a single link (Figure 2). The network has an obvious community structure, each community corresponding to a clique.



a) ***Determine the modularity Msingle of this natural partition, and the modularity Mpairs of the partition in which pairs of neighboring cliques are merged into a single community, as indicated by the dotted lines in Figure 2.***

- Each clique has $\frac{\mathrm{m(m\text{-}1)}}{2}$ links. So $\mathrm{N_c = m}$
- $\mathrm{M} = \sum_{c=1}^{n_c}\left[\frac{L_c}{L} - \left(\frac{k_c}{2L}\right)^2\right]$
- $M_{single} = \sum_{c=1}^{n_c}\left[\frac{m(m-1)/2}{\mathrm{n_c}\frac{\mathrm{m(m\text{-}1)}}{2}+\mathrm{n_c}} - \left(\frac{\mathrm{m(m-1)+2}}{\mathrm{2n_c}\frac{\mathrm{m(m\text{-}1)}}{2}+\mathrm{2n_c}}\right)^2\right] = \mathrm{n_c}\left[\frac{m(m-1)/2}{\mathrm{n_c}\frac{\mathrm{m(m\text{-}1)}}{2}+\mathrm{n_c}} - \left(\frac{\mathrm{m(m-1)+2}}{\mathrm{2n_c}\frac{\mathrm{m(m\text{-}1)}}{2}+\mathrm{2n_c}}\right)^2\right]$
    - $L_c = \frac{m(m-1)}{2}$
    - $L = n_c\frac{m(m-1)}{2} + \mathrm{n_c}$ (total links of $\mathrm{n_c}$ cliques plus total connected links $\mathrm{n_c}$, a ring shape has equal vertices and edges).
    - $k_c = \mathrm{m(m-1)+2}$ (a clique is a complete graph, so each vertex is connected to n-1 other vertices. The sum of all degrees in a clique is m(n-1), plus an additional 2 degrees because there are 2 connected links to it's 2 neighbors).
- $M_{\mathrm{pair}} = \sum_{c=1}^{n_c}\left[\frac{\mathrm{m(m\text{-}1)+1}}{\mathrm{n_c}\frac{\mathrm{m(m\text{-}1)}}{2}+\mathrm{n_c}} - \left(\frac{\mathrm{2m(m-1)+4}}{\mathrm{2n_c}\frac{\mathrm{m(m\text{-}1)}}{2}+\mathrm{2n_c}}\right)^2\right] = \frac{\mathrm{n_c}}{2}\left[\frac{\mathrm{m(m\text{-}1)+1}}{\mathrm{n_c}\frac{\mathrm{m(m\text{-}1)}}{2}+\mathrm{n_c}} - \left(\frac{\mathrm{2m(m-1)+4}}{\mathrm{2n_c}\frac{\mathrm{m(m\text{-}1)}}{2}+\mathrm{2n_c}}\right)^2\right]$
    - $L_c = m(m-1)+1$ (a pair of neighboring cliques has m(m-1) links and one link that connects them).
    - $k_c = 2m(m-1)+4$ (a pair of neighboring cliques has 2m(m-1) degrees, plus 2 degrees because there is 1 link connecting them, and an additional 2 degrees because each pair is connected to it's 2 neighbors with 2 links).

b) ***Show that only for $n_c < 2L$ will the modularity maximum predict the intuitively correct community partition, where $L = \frac{n_c m(m-1)}{2} + n_c$.***

- We have: $\mathrm{M} = \frac{\frac{\mathrm{L\text{-}n_c}}{\mathrm{n_c}}}{\mathrm{L}} - \left(\frac{\frac{\mathrm{2(L\text{-}n_c)}}{\mathrm{n_c}}+2}{\mathrm{2L}}\right)^2 = \frac{1}{\mathrm{n_c}} - \frac{1}{\mathrm{L}} - \frac{1}{\mathrm{n_c}^2}$
    - $\mathrm{L}_c = \frac{\mathrm{L\text{-}n_c}}{\mathrm{n_c}}$
    - k_c equals 2 multiply L_c plus 2 degrees because there are 2 connected links
- Because $n_c < 2L$: $\frac{1}{\mathrm{n_c}} > \frac{1}{\mathrm{2L}}$ and $\frac{1}{\mathrm{n_c}^2} > \frac{1}{\mathrm{4L^2}}$
- We have: $\frac{1}{\mathrm{n_c}} - \frac{1}{\mathrm{L}} > \frac{1}{\mathrm{2L}} - \frac{1}{\mathrm{L}} = -\frac{1}{\mathrm{L}}$

c) ***Discuss the consequences of violating the above inequality.***

Problem 4. Modularity Maximum

$$M = \sum_{c=1}^{n_c}[\frac{L_c}{L} - (\frac{k_c}{2L})^2]$$

Show that the maximum value of modularity M cannot exceed one.

- Because $\sum_{c=1}^{n_c} L_c \leq L$ (Total edges in all communities cannot exceed L)

$$M = \sum_{c=1}^{n_c}[\frac{L_c}{L} - (\frac{k_c}{2L})^2] = \sum_{c=1}^{n_c}\frac{L_c}{L} - \sum_{c=1}^{n_c}(\frac{k_c}{2L})^2 \leq \sum_{c=1}^{n_c}\frac{L_c}{L} = \frac{1}{L}\sum_{c=1}^{n_c} L_c \leq 1$$