Introduction To Louvain Method

Hsiu-Hsuan Yeh

2024-05-28

Resources

- Leskovec (2021)

Outline

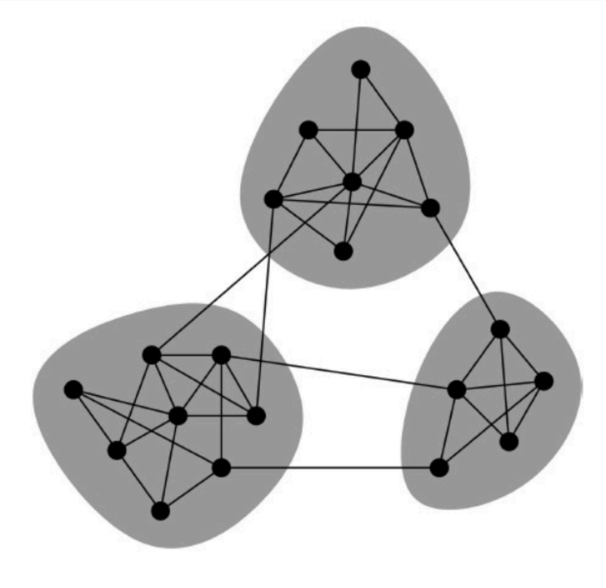
Modularity

Overview

Clauset et al. (2004) states that modularity measures when the division is a good one, in the sense that there are many edges within communities and only a few between them.

Modularity-based method including Blondel et al. (2008) evaluates the division by considering a hypothetical model. Given the division C, if the density of the connection within each community is higher than the hypothetical model, then C is somehow a relatively good division.

Overview



Formula

$$Q(C) = \frac{1}{2m}\Sigma_{i,j}(A_{ij} - \frac{k_i k_j}{2m})\delta(c_i, c_j)$$

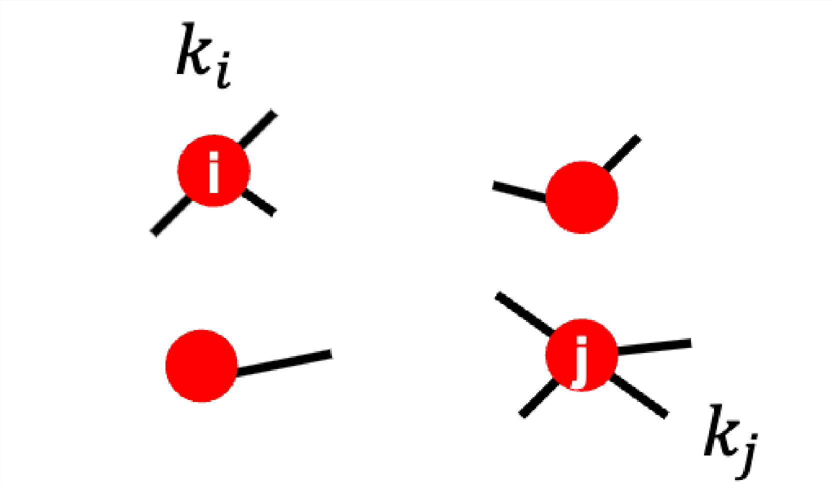
- C: the structure of the division, communities
- $A_{i,j}$: edge weight, can be a binary or real number
- k_i: $\Sigma_j A_{i,j}$, the total edge weight of a node i
- $2m$: $\Sigma_{i,j} A_{i,j}$, the total edge weight of a graph
- δ: indicating wheter the node i and j are in the same community

Hypothetical Model: Expected Edge Weight

$$Q(C) = \frac{1}{2m}\Sigma_{i,j}(A_{ij} - \frac{k_i k_j}{2m})\delta(c_i, c_j)$$

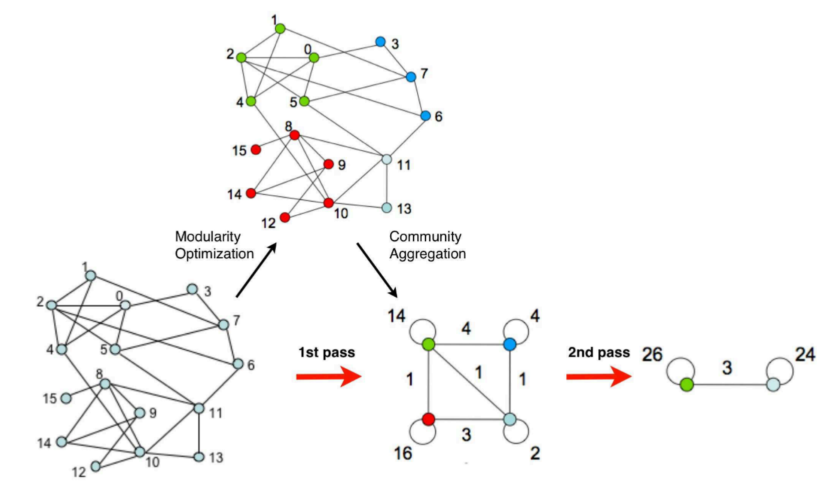
- $\frac{k_i k_j}{2m}$ is the expected edge weight for both node i and j
- assume each node remains its edge weight and the connection between nodes is random
  - view $2m$ as the number of all “available ports”
  - for node i, the edge weight assoicated to node j follows the binomial distribution, Binomial($k_i, \frac{k_j}{2m}$)
  - the expected edge weight, $E(\text{edge weight}_{i,j}) = k_i \frac{k_j}{2m}$

Hypothetical Model: Expected Edge Weight



Alogrithm

Algorithm



Modularity
Optimization
Community
Aggregation
1st pass
2nd pass
1
2
0
3
7
4
5
6
8
15
9
11
14
10
13
12
14
4
4
1
1
1
16
3
2
26
3
24

Algorithm

Modularity Optimization

- initialize each node in its own community
- for each node $i \in V$, compute the gain(ΔQ) in modularity by moving i to the community of each of its neighbors
- the node i will then be placed in the coummunity which has the hightest positive modularity gain(ΔQ)

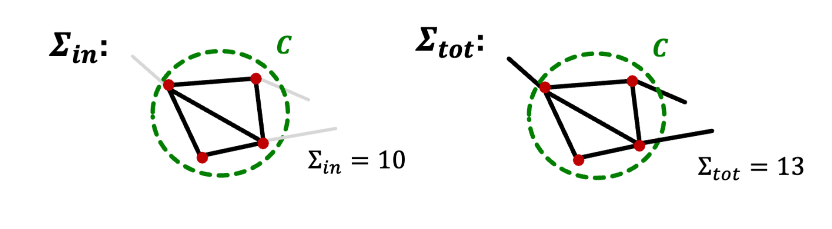
$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right]$$

Notice taht this is the change in Q when moving node i to a community. The change in Q when moving i out of a coummunity is similar.

Revision Of Modularity Formula

$$
\begin{aligned}
Q(C) &= \frac{1}{2m}\Sigma_{i,j}(A_{i,j} - \frac{k_i k_j}{2m})\delta(c_i, c_j) \\
&= \Sigma_{c \in C}\frac{\Sigma_{i,j\in c}A_{i,j}}{2m} - \frac{\Sigma_{i\in c}k_i \Sigma_{j\in c}k_j}{4m^2} \\
&= \Sigma_{c \in C}\frac{\Sigma_{in,c}}{2m} - \frac{\Sigma_{tot,c}\Sigma_{tot,c}}{4m^2} \\
&= \Sigma_{c \in C} \underbrace{\frac{\Sigma_{in,c}}{2m}}_{\text{edge weight within } c} - (\underbrace{\frac{\Sigma_{tot,c}}{2m}}_{\text{total edge weight of } c})^2
\end{aligned}
$$

Some Notations



- $\Sigma_{in,c} = \Sigma_{i,j \in c} A_{i,j}$
- $\Sigma_{tot,c} = \Sigma_{i \in c} k_i = \Sigma_{i \in c} \Sigma_j A_{i,j}$

More Notaions

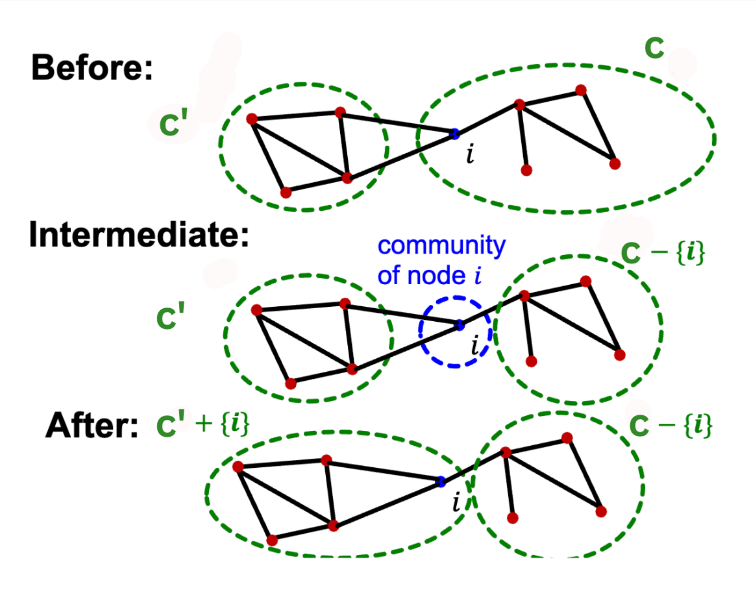
Considering a node i moving from community c to c', there are two types of edge weights related to i.

- $k_i = k_{i,c'} + k_{i,c}$
- associated to community c': $k_{i,c'}$
- associated to community c: $k_{i,c}$

What we want to know is the change in modularity, $\Delta Q(C)$ for node i moving from c to c'. It consists of two components: moving into c' and moving out of c.

$\Delta Q(C) = \Delta Q(c') + \Delta Q(c)$

Change In Modularity



Moving Into c'

- $\Delta Q(c') = Q_{i \to c'}(c') - (Q(c') + Q(i))$
- $Q(c')$: modularity of c' before merging with node i
  - $\frac{\Sigma_{in,c'}}{2m} - (\frac{\Sigma_{tot,c'}}{2m})^2$
- $Q(i)$: modularity of a community with a unique member i
  - $0 - (\frac{k_i}{2m})^2 = -(\frac{k_i}{2m})^2$
- $Q_{i \to c'}(c')$: modularity of c' after merging with node i
  - $\frac{\Sigma_{in,c'} + 2k_{i,c'}}{2m} - (\frac{\Sigma_{tot,c'} + k_i}{2m})^2$

Moving Into c'

$$
\begin{aligned}
\Delta Q(c') &= Q_{i\rightarrow c'}(c') - (Q(c') + Q(i)) \\
&= \underbrace{\frac{\Sigma_{in,c'} + 2k_{i,c'}}{2m} - (\frac{\Sigma_{tot,c'} + k_i}{2m})^2}_{Q_{i\rightarrow c'(c')}} \\
&- [\underbrace{\frac{\Sigma_{in,c'}}{2m} - (\frac{\Sigma_{tot,c'}}{2m})^2}_{Q(c')} \underbrace{-(\frac{k_i}{2m})^2}_{Q(i)}] \\
&= \frac{2k_{i,c'}}{2m} - \frac{2k_i \Sigma_{tot,c'}}{4m^2}
\end{aligned}
$$

Moving Out of c

- $\Delta Q(c) = (Q_{c \to i}(c) + Q(i)) - Q(c)$
- $Q(c)$: modularity of c before node i moving out
  - $\frac{\Sigma_{in,c}}{2m} - \left(\frac{\Sigma_{tot,c}}{2m}\right)^2$
- $Q_{c \to i}(c)$: modularity of c after node i moving out
  - $\frac{\Sigma_{in,c} - 2k_{i,c}}{2m} - \left(\frac{\Sigma_{tot,c} - k_i}{2m}\right)^2$

Moving Out of c

$$
\begin{aligned}
\Delta Q(c) &= (Q_{c\to i}(c) + Q(i)) - Q(c) \\
&= [\underbrace{\frac{\Sigma_{in,c} - 2k_{i,c}}{2m} - (\frac{\Sigma_{tot,c} - k_i}{2m})^2}_{Q_{c\to i(c)}} \underbrace{-(\frac{k_i}{2m})^2}_{Q(i)}] \\
&- [\underbrace{\frac{\Sigma_{in,c}}{2m} - (\frac{\Sigma_{tot,c}}{2m})^2}_{Q(c)}] \\
&= \frac{-2k_{i,c}}{2m} + \frac{2k_i\Sigma_{tot,c}}{4m^2} - 2(\frac{k_i}{2m})^2
\end{aligned}
$$

Modularity Gain

$$
\begin{aligned}
\Delta Q(C) &= \Delta Q(c') + \Delta Q(c) \\
&= \underbrace{\frac{2k_{i,c'}}{2m} - \frac{2k_i\Sigma_{tot,c'}}{4m^2}}_{\Delta Q(c')} + \underbrace{\frac{-2k_{i,c}}{2m} + \frac{2k_i\Sigma_{tot,c}}{4m^2} - 2(\frac{k_i}{2m})^2}_{\Delta Q(c)} \\
&= \frac{2(k_{i,c'} - k_{i,c})}{2m} + \frac{2k_i((\Sigma_{tot,c} - k_i) - \Sigma_{tot,c'})}{4m^2}
\end{aligned}
$$

The intuition for positive modularity gain movement:

- for node i, there are more connection in community c' comapred to c, $k_{i,c'} > k_{i,c}$
- the size of c' is smaller than c, $\Sigma_{tot,c} - k_i > \Sigma_{tot,c'}$

Pros and Cons

Pros

- comprehensive
- efficient - the height of dendrogram is usually short

Cons

- high extra-space complexity: $O(|V|^2)$
- local optimization
- potential instability - the order of nodes for modularity optimization matters

Application to large networks

	Karate	Arxiv	Internet	Web nd.edu	Phone	Web uk-2005	Web WebBase 2001
Nodes/links	34/77	9k/24k	70k/351k	325k/1M	2.6M/6.3M	39M/783M	118M/1B
CNM	.38/0s	.772/3.6s	.692/799s	.927/5034s	-/-	-/-	-/-
PL	.42/0s	.757/3.3s	.729/575s	.895/6666s	-/-	-/-	-/-
WT	.42/0s	.761/0.7s	.667/62s	.898/248s	.56/464s	-/-	-/-
Our algorithm	.42/0s	.813/0s	.781/1s	.935/3s	.769/134s	.979/738s	.984/152mn

References

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, 2008, *2008* (10), P10008.

Clauset, Aaron, Mark EJ Newman, and Cristopher Moore, "Finding community structure in very large networks," *Physical review E*, 2004, *70* (6), 066111.

Leskovec, Jure, "CS224W: Community Structure in Networks," `https://snap.stanford.edu/class/cs224w-2021/slides/14-communities.pdf` 2021. Accessed: 2024-05-23.