Community Detection in Signed Bipartite Networks

Rishi Bharadwaj

March 2024

1 Modularity

Modularity is a measure of how good a particular partitioning is for a given graph. It involves finding the difference between the number of edges between two nodes present in the same community and the number of edges that would be present between those nodes in a null or random model. In unsigned bipartite graphs Barber's bipartite modularity is generally used [1].

$$Q_B = \frac{1}{m} \sum_{i=1} \sum_{j=1} \left(B_{ij} - \frac{k_i d_j}{m} \right) \delta(c_{u_i}, c_{v_j})]$$

where m is the number of edges, $B_{ij} \in R^{N \times M}$ indicates the connections between U and V, which are the two sets of nodes of N and M features respectively, k_i and d_j are the degrees of node u_i and v_j, respectively, and c_{u_i} and c_{v_j} indicate the community index of u_i and v_j, respectively.δ is 1 if both nodes belong to the same community, or else it is 0.

This can be extended to weighted graphs.

$$Q_B = \frac{1}{w} \sum_{i=1} \sum_{j=1} \left(W_{ij} - \frac{w_i w_j}{w} \right) \delta(c_{u_i}, c_{v_j})]$$

where w is the sum of the weights of all the edges, $W_{ij} \in R^{N \times M}$ indicates the weight of the connections between U and V, w_i and w_j are the sum of the edge weights connected to node u_i and v_j, respectively, and c_{u_i} and c_{v_j} indicate the community index of u_i and v_j, respectively.

To extend modularity to signed bipartite graphs we can follow a similar approach to [2] and separate the positive and negative edge weight contributions.

$$w_{ij} = w_{ij}^+ - w_{ij}^-$$

The positive and negative strengths are given by

$$w^+ = \frac{1}{2} \sum_{i=1} \sum_{j=1} w_{ij}^+$$

$$w^- = \frac{1}{2}\sum_{i=1}\sum_{j=1} w_{ij}^-$$

Finding modularity separately for the positive and negative edges we get

$$Q^+ = \frac{1}{w^+}\sum_i \sum_j \left(w_{ij}^+ - \frac{w_i^+ w_j^+}{w^+} \right) \delta(c_i, c_j)$$

$$Q^- = \frac{1}{w^-}\sum_i \sum_j \left(w_{ij}^- - \frac{w_i^- w_j^-}{w^-} \right) \delta(c_i, c_j)$$

Combining these and weighing them in proportion to the total strengths of the positive and negative weights we get

$$Q = \frac{w^+}{w^+ + w^-} Q^+ + \frac{w^-}{w^+ + w^-} Q^-$$

2 Balance Theory

Balance theory is based on the principle that "The enemy of my friend is my enemy" and "The friend of my friend is my friend". This is represented in signed networks through the use of positive and negative edge weights. If there are an even number of negative edges between two nodes the two nodes can be considered friends and the path is balanced. If there are an odd number of negative edges the two nodes can be considered enemies and the path is unbalanced. In unipartite graphs triads formed by three nodes demonstrate this principle. In bipartite graphs this principle can be extended to signed butterflies. A butterfly is balanced when it contains an even number of negative edges.

3 Message Passing

Message passing in signed graphs must incorporate balance theory. The unipartite approach which does not incorporate the signs of edges is not optimal. Instead, as proposed in [3], for each node we can learn separate representations of the node's friends and enemies. For each node we can define $B_k(l)$ and $U_k(l)$ as the sets of nodes that are balanced and unbalanced with respect to some node u_k, where l is the number of hops from the current node. They can be defined recursively as follows.

$$\text{When } l = 1, \quad B_i(1) = \{u_j | u_j \in N_i^+\}; \quad U_i(1) = \{u_j | u_j \in N_i^-\}.$$

$$\text{For } l > 1, \quad B_i(l+1) = \{u_j | u_k \in B_i(l) \text{ and } u_j \in N_k^+\} \cup \{u_j | u_k \in U_i(l) \text{ and } u_j \in N_k^-\}$$

$$U_i(l+1) = \{u_j | u_k \in U_i(l) \text{ and } u_j \in N_k^+\} \cup \{u_j | u_k \in B_i(l) \text{ and } u_j \in N_k^-\}.$$

Where N_k^+ is the set of nodes connected to u_k through positive links and N_k^- is the set of nodes connected through negative links. For a given node u_k and a given length l all members of $B(l)$ and $U(l)$ will belong only to V or U. If we start from a node u_k in U all $B(l_{\in 2k})$ and all $B(l_{\in 2k+1})$ will comprise entirely of nodes from U and V respectively, and vice-versa if we start from a node in V.

As the graph is bipartite we also need to consider the aggregation of information from two sets of nodes which may have different features. Thus we propose the use of four different linear layers W_u, W_v, W^B,W^U to transform feature dimensions into common dimensions and to maintain separate representations of the balanced and unbalanced sets. For the first aggregation layer:

$$h_u^{B(1)} = \sigma(W^{B(1)} \cdot AGG[(h_v^{(0)}, \forall h_v \in N_u^+), h_u^{(0)}])$$

$$h_u^{U(1)} = \sigma(W^{U(1)} \cdot AGG[(h_v^{(0)}, \forall h_v \in N_u^-), h_u^{(0)}])$$

$$h_v^{B(1)} = \sigma(W^{B(1)} \cdot AGG[(h_u^{(0)}, \forall h_u \in N_u^+), h_v^{(0)}])$$

$$h_v^{U(1)} = \sigma(W^{U(1)} \cdot AGG[(h_u^{(0)}, \forall h_u \in N_v^-), h_v^{(0)}])$$

where $h_u^{(0)}$ and $h_v^{(0)}$ represent the initial node features of some node in U and V respectively. N_u^+ and N_u^- are the sets of nodes in V connected to a node in U through positive and negative links respectively. N_v^+ and N_v^- are the sets of nodes in U connected to a node in V through positive and negative links respectively. The aggregation feature will make use of W_u and W_v to maintain common dimensionality of both types of nodes. For example, the aggregation can be the mean of all positive one hop neighbours concatenated with the representation of the anchor node.

$$AGG = W_v \cdot [\frac{\sum h_v}{N_u^+}], W_u \cdot h_u$$

For subsequent layers:

If $l \in 2k$

$$h_u^{B(l)} = \sigma(W^{B(l)} \cdot AGG[(h_u^{B(l-1)}, \forall h_u \in N_u^+), (h_u^{U(l-1)}, \forall h_u \in N_u^-), h_u^{l-1}])$$

$$h_u^{U(l)} = \sigma(W^{U(l)} \cdot AGG[(h_u^{U(l-1)}, \forall h_u \in N_u^+), (h_u^{B(l-1)}, \forall h_u \in N_u^-), h_u^{l-1}])$$

If $l \in 2k+1$

$$h_u^{B(l)} = \sigma(W^{B(l)} \cdot AGG[(h_v^{B(l-1)}, \forall h_v \in N_u^+), (h_v^{U(l-1)}, \forall h_v \in N_u^-), h_u^{l-1}])$$

$$h_u^{U(l)} = \sigma(W^{U(l)} \cdot AGG[(h_v^{U(l-1)}, \forall h_v \in N_u^+), (h_v^{B(l-1)}, \forall h_v \in N_u^-), h_u^{l-1}])$$

4 Loss Function

After learning the two representations of each node we can get the assumed cluster assignment of each node by passing the representations to a softmax layer, as done in [4]. We can define $C_U \in R^{N\times K}$ and $C_V \in R^{M\times K}$ as two node-community assignment matrices for U and V, respectively. Each element of $C_U = \{c_{ik}\}$ and $C_V = \{c_{jk}\}$ can take binary values 0 or 1, which indicates whether or not this particular node i or j is assigned to community k. We can get C_U and C_V from the representations using a softmax layer after a linear layer, i.e, $C_U = \text{softmax}(\text{LL}(H_u))$ and $C_V = \text{softmax}(\text{LL}(H_v))$, where LL denotes a fully-connected linear layer with weight parameter W_L of proper dimension.

Next we can define a modularity matrix $B \in R^{N\times M}$ such that each entry

$$q_{ij} = q_{ij}^+ + q_{ij}^-$$

where

$$q_{ij}^+ = \left(w_{ij}^+ - \frac{w_i^+ w_j^+}{w^+} \right)$$

and

$$q_{ij}^- = \left(w_{ij}^- - \frac{w_i^- w_j^-}{w^-} \right)$$

Then modularity can be defined as

$$Q = \frac{1}{w^+ + w^-} Tr(C_U^T B C_V)$$

Adding Frobenius regularization terms to avoid trivial solutions, we get the following loss function.

$$L = \frac{-1}{w^+ + w^-} Tr(C_U^T B C_V) + \sqrt{\frac{K}{N}} \left\| \sum_i C_i^U \right\|_F + \sqrt{\frac{K}{M}} \left\| \sum_i C_i^V \right\|_F - 2$$

References

[1] Michael J Barber. 2007. Modularity and community detection in bipartite networks. Physical Review E, Vol. 76, 6 (2007), 066102.

[2] Gomez, Sergio & Jensen, Pablo & Arenas, Alex. (2009). Analysis of community structure in networks of correlated data. Physical review. E, Statistical, nonlinear, and soft matter physics. 80. 016114. 10.1103/PhysRevE.80.016114.

[3] T. Derr, Y. Ma and J. Tang, "Signed Graph Convolutional Networks," in 2018 IEEE International Conference on Data Mining (ICDM), Singapore, Singapore, 2018 pp. 929-934. doi: 10.1109/ICDM.2018.00113

[4] Cangqi Zhou, Yuxiang Wang, Jing Zhang, Jiqiong Jiang, and Dianming Hu. End-to-end modularity-based community co-partition in bipartite networks. In Proceedings of the 31st CIKM, pages 2711–2720. ACM, 2022.