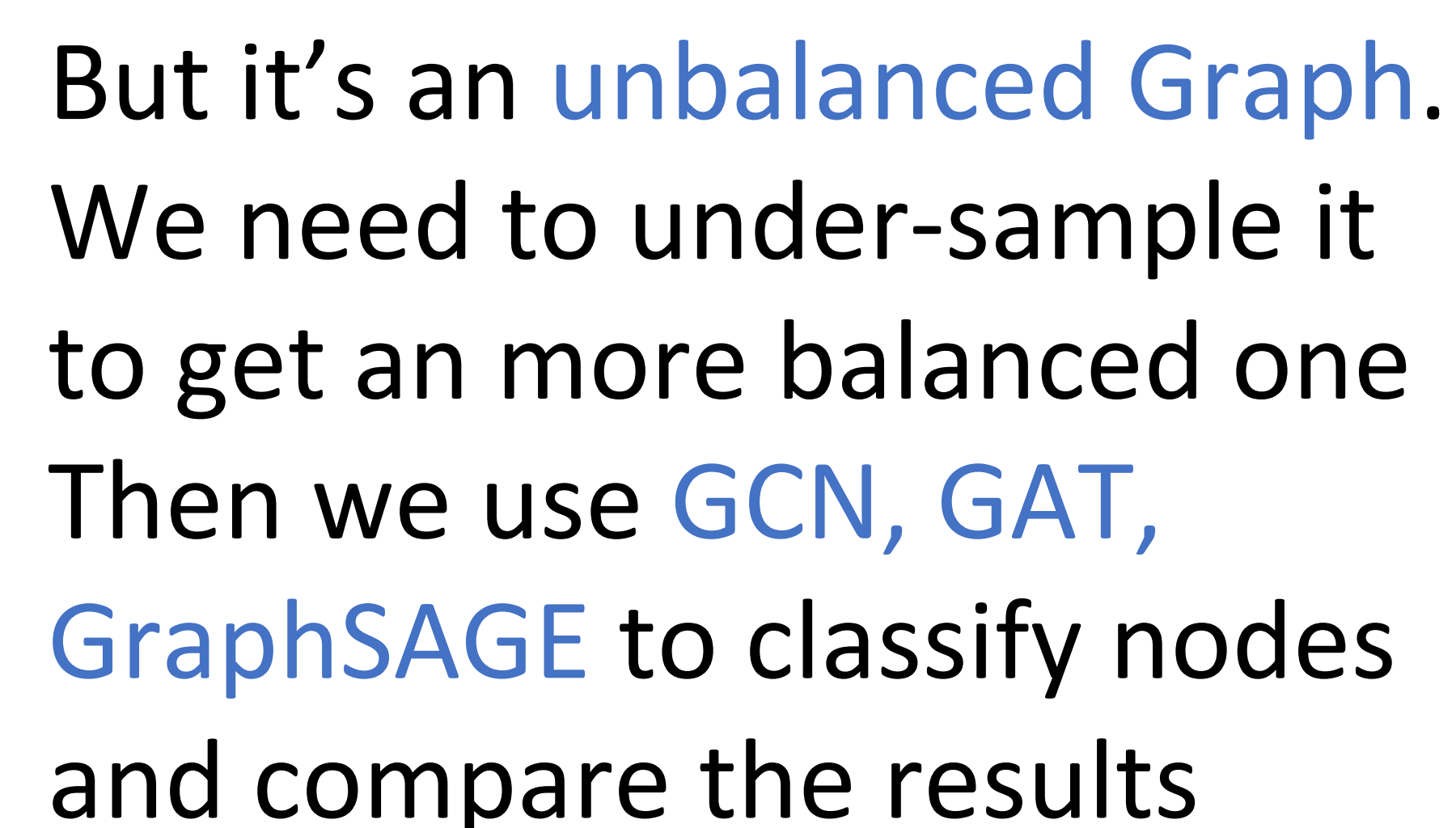




Code Available at : <https://github.com/TianshuoLi/DataMining>

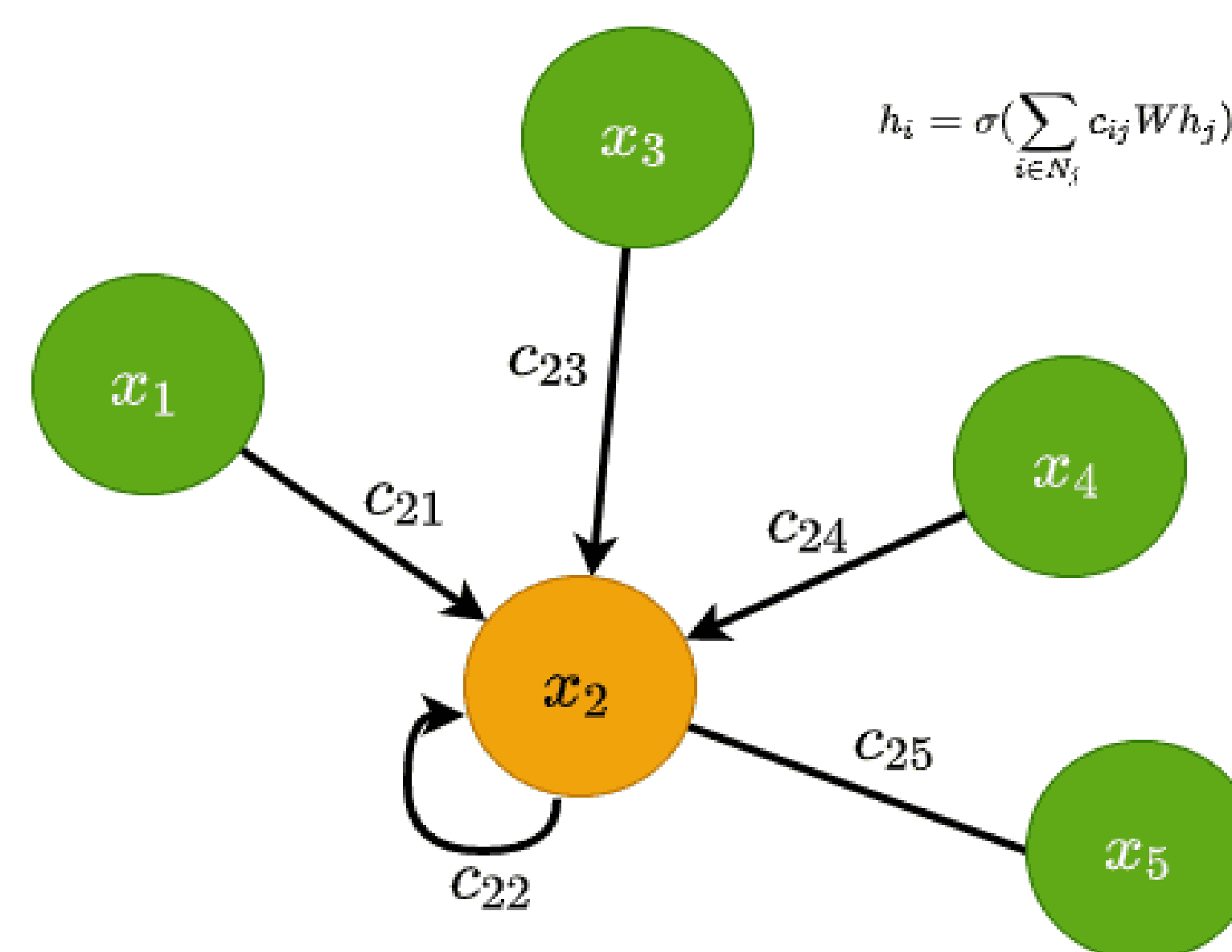
PPI file be like. if it is causing Asthma/ Schizophrenia, label 1 to it and 0 otherwise



If H is the feature matrix and W is the trainable weight matrix, From a **node-wise** perspective, the update rule can be written as:

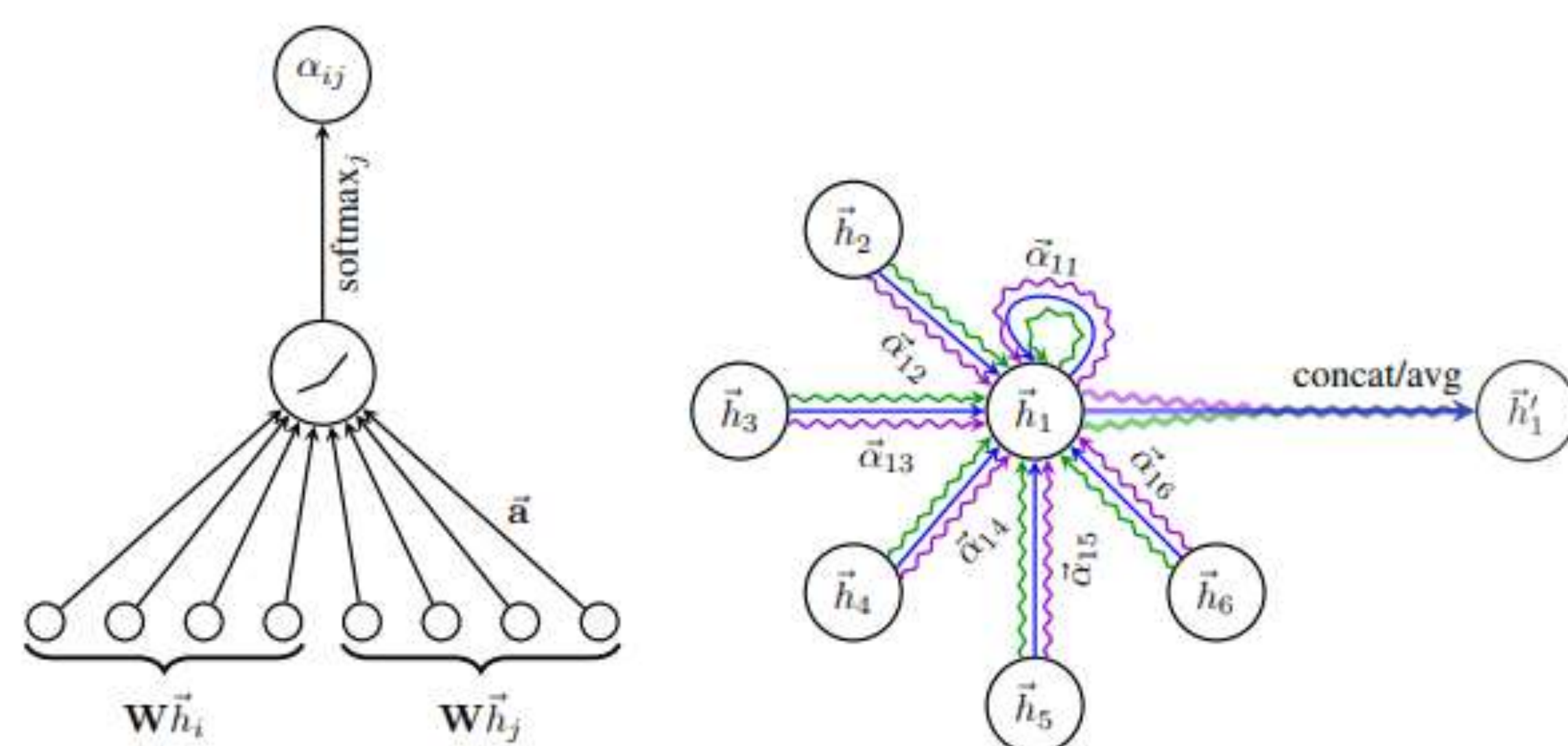
The coefficient $c_{ij} = \frac{1}{\sqrt{N_i N_j}}$

Where N_i is the size of the node's neighbor



In GAT, we considering the coefficient to be a **learnable attention** mechanism. Mathematically:

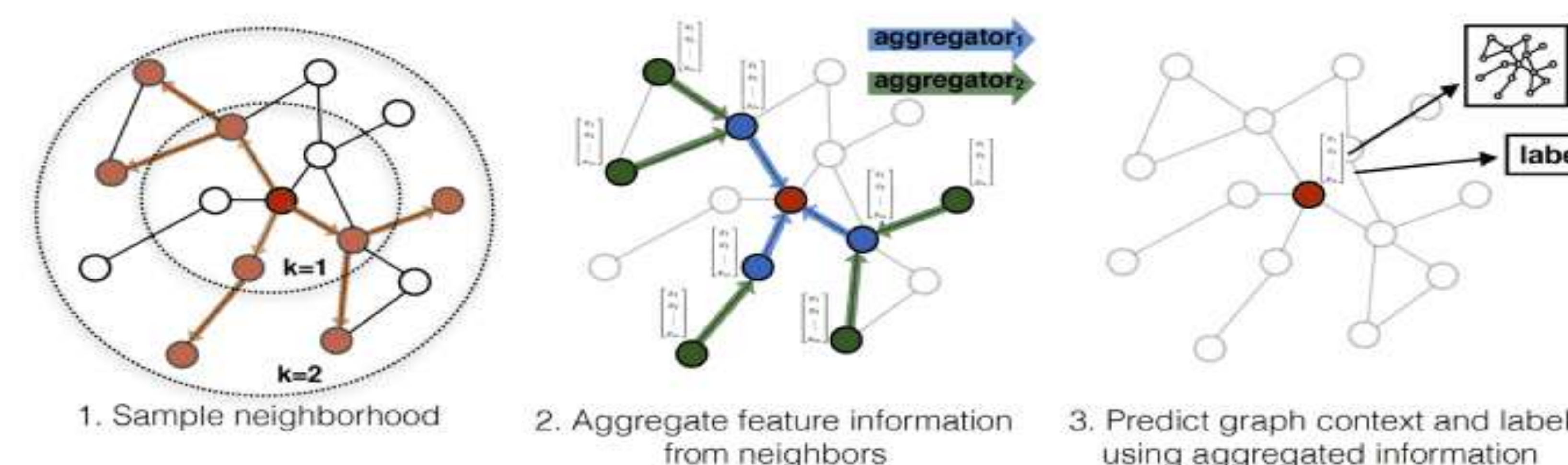
$$a_{ij} = \frac{\exp(a_{ij})}{\sum_{k \in N_i} \exp(a_{ik})}$$



GraphSAGE (Graph Sample and aggregate)

GraphSAGE proposes the following framework:

1. **Sample** uniformly a set of nodes from the neighbors
2. **Aggregate** the feature information from sampled neighbors
3. Based on the aggregation, perform node classification



Experimental results

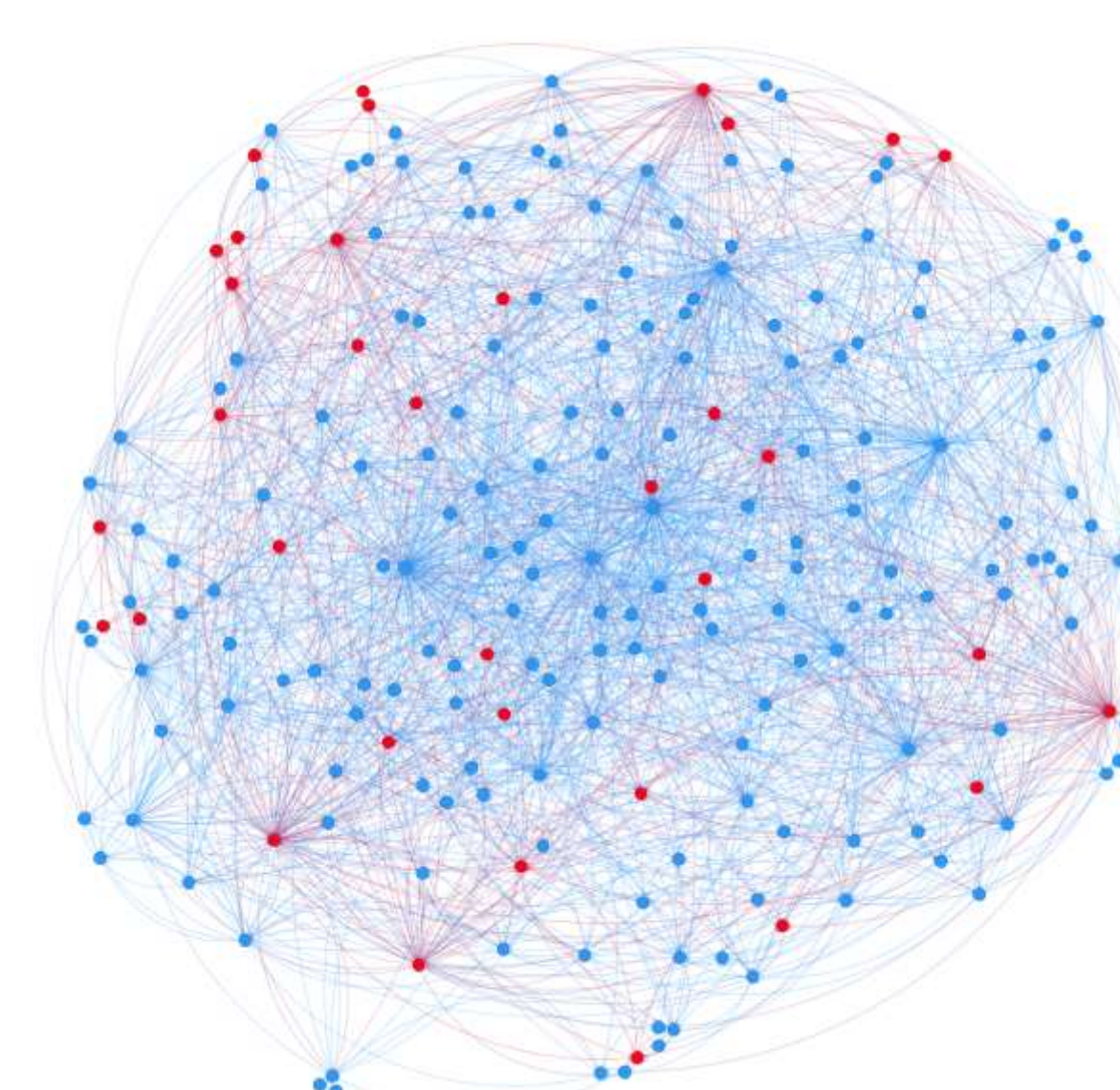
- Results on Asthma

		Accuracy	Precision	F1-score	recall	TP	TN	FP	FN
GCN	train	0.657	0.333	0.407	0.522	36	165	72	33
	test	0.574	0.417	0.508	0.652	30	48	42	16
GAT	train	0.754	0.425	0.312	0.246	17	213	23	52
	test	0.74	0.619	0.433	0.333	13	84	8	26
SAGE	train	0.744	0.235	0.093	0.058	4	223	13	65
	test	0.71	0.8	0.182	0.103	4	91	1	35

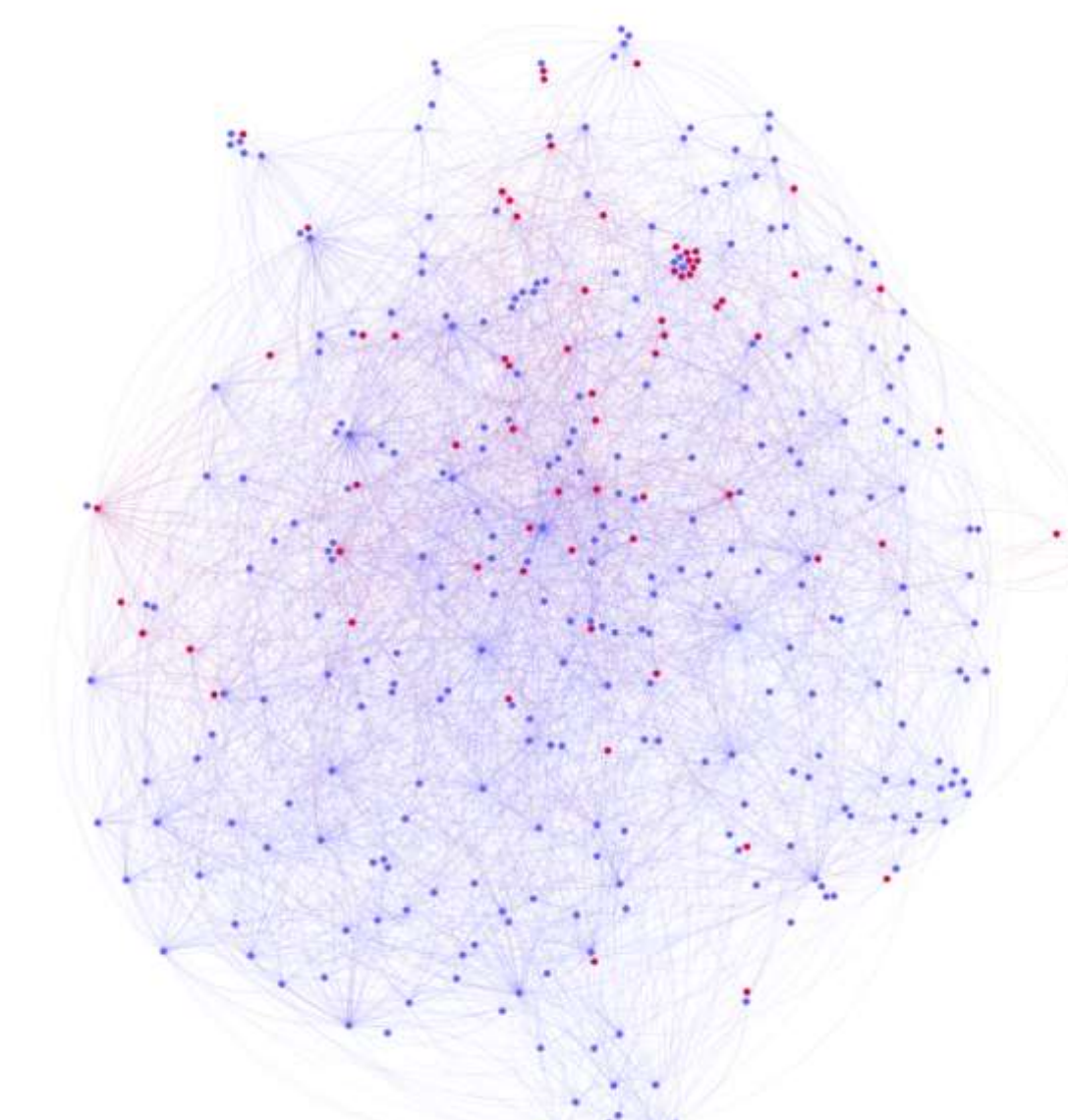
- Results on Schizophrenia

		Accuracy	Precision	F1-score	recall	TP	TN	FP	FN
GCN	train	0.691	0.749	0.814	0.89	626	17	210	77
	test	0.782	0.782	0.878	1	312	0	87	0
GAT	train	0.756	0.756	0.861	1	703	0	227	0
	test	0.782	0.782	0.878	1	312	0	87	0
SAGE	train	0.732	0.757	0.843	0.950	668	13	214	35
	test	0.781	0.792	0.875	0.977	305	7	80	7

- Selected results Visualization



GCN on Asthma



GAT on Asthma