Print

Featurizing a Bipartite Graph

To featurize a bipartite graph, we can first consider the ego network (or egonet) for each company, which consists of the company and its direct resources. Features can then be created representing the number of links to fraudulent resources (dF), the number of links to non-fraudulent resources (dNF), and the relative number of links to fraudulent resources, which is dF divided by (dF plus dNF).

Another interesting feature is based on the fraud score of each of the resources connected to the company. The fraud score of a resource can be computed as the ratio of the number of fraudulent companies connected to the resource and the total number of companies connected to the resource. The fraud scores of the resources connected to a company can then be summarized by calculating the minimum, maximum, or average. In the previous example, the resources were analyzed independently. However, resources might also be connected. If multiple resources that are shifting toward a company come from the same fraudulent company, then this substantially increases the fraud risk.

One way of capturing the relationships is by defining triangles in the egonet. In this diagram, resources 18 and 19 were already connected to the fraudulent company A. Both are now connected to company B, which seriously increases the fraud risk of company B. To appropriately model this behavior, we will connect resources 18 and 19 with an edge, which creates a triangle with company B. After the triangles have been added to the egonet, they can be featurized by counting the number of fraudulent and non-fraudulent triangles. A weight can also be assigned to each triangle based on the number of times the resources are associated with the same fraudulent company.

Close