

## Community Mining: Techniques

Graph partitioning techniques can also be used for community mining. The aim here is to partition a social network into communities. The most popular approach is iterative bisection, where the goal is to find the best division of the complete network into two communities, or the best two-group cut, and then further subdivide those two until we reach the required number of communities. A cut is defined as the set of edges that has only one node in each community, or in other words, the between-community edges. In the example, the cut consists of the edges A-D and C-F. The goal of iterative bisection is to partition the graph according to the cut that has the minimum size.

Various metrics can be used to evaluate the size of a cut, such as the min cut, ratio cut, and min-max cut. Let's discuss these in more detail.

The min-cut metric chooses the partitioning so that the sum of the weights of the between-community edges or the cut is minimal. In the example, the min cut equals 0.3. A key disadvantage of this metric is that it does not consider the size of the communities. Hence, it might result in the separation of one node from all of the others, which is clearly not desirable.

The ratio-cut metric takes into account the size of the communities. It is calculated as min cut (A,B) times (1 divided by the size of A plus 1 divided by the size of B). In this case, it is equal to 0.3 times (1 divided by 3 plus 1 divided by 4), or 0.175. The obtained communities will typically be more balanced. A shortcoming of the min-cut and ratio-cut metrics is that the within-community edges are not taken into account.

The min-max cut takes both the within- and between-community edges into consideration. It is defined as follows: min cut (A,B) times (1 divided by cut (A,A) plus 1 divided by cut (B,B)), whereby cut (A,A) and cut (B,B) represent the sum of the weights of the within-community edges for communities A and B, respectively. In the example, the value becomes 0.3 times (1 divided by (0.7 plus 0.8 plus 0.9) plus 1 divided by (0.6 plus 0.6 plus 0.8 plus 0.8 plus 0.7)), or 0.211.

---

### *Social Network Analytics*

Copyright © 2019 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close