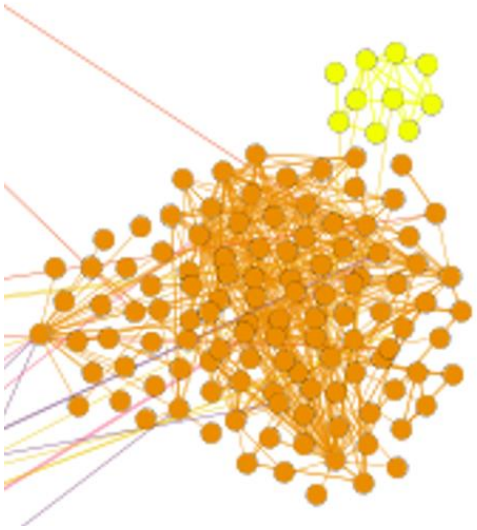
Clustering Methods

Exercises 6/7

1. Internal inverse weight (IIW) was shown to balance cluster sizes. Show by example why this happens.

2. Three clusters were found around central nodes of three football club Twitter accounts: Manchester United, Liverpool, and Arsenal. Assume that no user follows more than one of these accounts, and that users within the same club follow each other with 10% probability, and users of another club with 2% probability. Now suppose that a new user joins and follows all three clubs. Which cluster will the user be assigned to by (a) Conductance, (b) IIW, (c) MIW?

3. Sub-cluster in Iceland was often clustered to different country (Sweden). The problem roots in the random initialization, weak connectivity and tendency of users to join clusters of their neighbors. If one user in this sub-cluster joins Sweden, the others are likely to follow. Can you explain why it is difficult for this sub-cluster to change cluster later?



4. Four network strength measures were used for clustering ego networks. Which of these measures would you think is most sensible? Or how would you measure the strengths of the network?

5. Pre-processing aims at unifying the scales of the different attributes. What will happen if one of the attributes has very low variance? What if the variance is very large? How would you do this scale normalization?