Native Communities in Social Networks

Igor Kanovsky[1].

(1) The Max Stern Yezreel Valley College, Emek Yezreel, Israel, igork@yvc.ac.il.

A lot of methods for clusters discovery in complex networks were developed in last decade. The main approach is based on cluster definition as a subnetwork with bigger than average link density. From the other hand in social networks each actor can define the communities he/she belongs to without information about a network as a whole. Native community is a group of actors which formed the community by creating inside community links. Native community may be detected by near neighbor actors' attributes analysis. An actor can belong to more than one native community.

Let *commonity c(i,j)* of two actors *i* and *j* be a fraction of common neighbors:

$$c(i, j) = \frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|}$$

where $N(i)$ is the neighborhood of the actor *i*. A link, which connect two actors with big commonity $c(i,j) > c_0$, is "for sure" an inner link of native community. On Fig. (A,B), (B,C), (D,E), (E,F) are "for sure" inner links, so these couples belongs to the same community. Two actors k,l having inner links to the same in between actor are "for sure" in the same community if there commonity is big enough $c(k,l) > c_1$. On Fig. A, C belong to different communities, and B belongs to both overlapping communities. Contrary, D, F belong to the same community. Threshold values c_0, c_1 have to be chosen in depends on links density of the network, and usually they are network's clustering coefficient value, and usually they are network's clustering coefficient value.
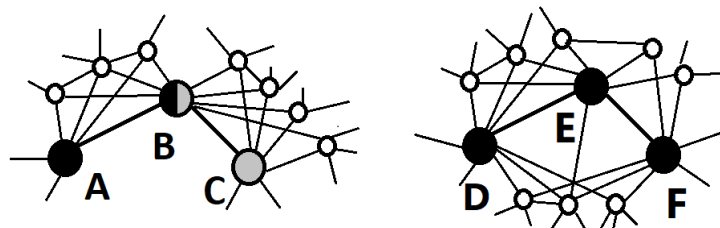

Figure 1: Network fragments. Depicted three nodes and all their common neighbors.
c(A,B)=3/7, c(B,C)=4/10, c(A,C)=1/13, c(D,E)=4/11, c(E,F)=3/11, c(D,F)=4/13.

Based on this approach we developed communities' detection algorithm. Starting with arbitrary chosen actor, by links commonity calculation in iterative manner, all overlapping communities are detected. The algorithm was applied to different real world social networks and to big networks with communities generated for testing. Effectiveness and efficiency of the algorithm were investigated. The algorithm is very effective for social networks with big clustering coefficient. Advantages of the approach are locality (no need of information about whole network to detect a community) and stability (after changing some amount of links the same communities still are recognizable).