Preferential Placement for Community Structure Formation

Aleksandr Dorodnykh
Moscow Institute of Physics and Technology, Moscow, Russia
e-mail: dorodnykh@phystech.edu

Liudmila Ostroumova Prokhorenkova
Moscow Institute of Physics and Technology, Moscow, Russia
e-mail: ostroumova-la@yandex.ru

Egor Samosvat
Moscow Institute of Physics and Technology, Moscow, Russia
e-mail: samosvategor@gmail.com
Moscow Institute of Physics and Technology, Moscow, Russia

Various models have been recently proposed to reflect and predict different properties of complex networks. However, the community structure, which is one of the most important properties, is not well studied and modeled.

Several empirical studies have shown that community structure of different real-world networks has some typical properties. In particular, it was observed that the cumulative community size distribution obeys a power law with some parameter λ. We suggest a principle called "preferential placement", which allows to model a realistic community structure. Our approach is based on the idea that vertices can be embedded in a multidimensional space of latent features. The vertices appear one by one and their positions are defined according to preferential placement: new vertices are more likely to fall into already dense regions. Namely, each new vertex chooses a "parent" uniformly at random from the set of existing vertices and then it is placed according to some distribution Ξ centered at this parent vertex. We argue that in order to obtain a realistic clustering structure one should take a heavy tailed distribution Ξ. In this case, new vertices usually appear in the dense regions, close to some previously added vertices; however, due to the heavy tail of Ξ, from time to time we get outliers, which originate new clusters.

We provide an extensive empirical analysis of the obtained structure as well as some theoretical heuristics. In particular, our analysis shows that the distribution of cluster sizes follows a power law, as it is observed in real-world complex networks.

Note that, given the embeddings for all vertices, one can construct a graph using one of many well-known approaches. The basic idea behind such approaches is that we want to increase the probability of connecting two vertices if they have similar latent features.

More details on this topic can be found in [1].

References

1. A. Dorodnykh, L. Ostroumova Prokhorenkova, E. Samosvat, "Preferential Placement for Community Structure Formation", *WAW 2017, LNCS 10519*, 1–15 (2017).