

Localization of Hubs in Modular Networks

Z. Ghalmane^{1,3}, C. Cherifi², H. Cherifi¹, M. El Hassouni³⁽¹⁾ University of Burgundy, Dijon, France. (2) DISP Lab, University of Lyon 2, Lyon, France. (3) LRIT, Rabat IT center, Mohammed V University, Rabat, Morocco.

In complex networks, the degree distribution of the nodes is known to be non-homogeneous with a heavy tail. Consequently, a small set of nodes (called hubs) are highly connected while the vast majority share few connections with their neighbors. The community structure is another main topological feature of many real-world networks. In these networks, the nodes shared by more than one community are called overlapping nodes. They play an important role in the network dynamics due to their ability to reach multiple communities¹. In this work, our goal is to characterize the relationship between the overlapping nodes and the hubs. Indeed, we suspect that hubs are in the neighborhood of the overlapping nodes. In order to investigate the ubiquity of this property, we perform series of experiments on multiple real-world networks from various origins. The aim of these experiments is to compare the set of neighbors of the overlapping nodes with the set of hubs. In order to define both sets, the overlapping community structure of the real-world networks is uncovered using an overlapping community detection algorithm. The overlapping nodes and the set of their neighbors are then extracted. Note that if n is the size of the neighbors of the overlapping nodes, the set of hubs is formed using the top n nodes of the network. This choice is motivated by the fact that some similarity measures need to be computed on sets of the same size. We compute classical measures such as the proportion of common nodes, the Jaccard Index, the Rank-biased overlap, the correlation measures between the two sets (Pearson and Spearman) in order to investigate their similarities. We also study their degree distribution and compare the subgraph made of the overlapping nodes and the hubs with the original network. Table 1 illustrates partial results of these evaluation for 10 empirical networks of different sizes using two different community detections algorithms to uncover the overlapping nodes (SLPA and LFME).

Network	N	$on(\%)$	$S(\%)$	$A_n(\%)$		$A_t(\%)$		$r(p=0.5)$		$r(p=0.98)$		ρ		ρ_s	
				SLPA	LFME	SLPA	LFME	SLPA	LFME	SLPA	LFME	SLPA	LFME	SLPA	LFME
Karate club	34	15.1	23.52	77.6	80	100	100	0.997	0.805	0.923	0.772	0.991	0.919	0.98	0.987
Dolphins	62	24	35.48	77.89	75	83.33	80.17	0.989	0.998	0.956	0.914	0.973	0.972	0.944	0.972
Les Miserables	77	2.5	44.15	81.06	76.47	100	98	0.999	0.997	0.91	0.852	0.979	0.979	0.972	0.983
Game of thrones	107	7.4	33.64	67.73	73.68	100	97.92	0.999	0.999	0.978	0.979	0.99	0.997	0.99	0.996
College football	115	7.82	33.17	65.21	75.86	54.54	59.3	0.996	0.999	0.834	0.921	0.804	0.85	0.81	0.86
ego-Facebook	4039	3.1	17.5	59.25	71.83	50.38	57.14	0.998	0.999	0.844	0.998	0.985	0.99	0.98	0.99
ca-GrQc	5242	26.43	33.19	64.04	64.01	65.12	74.69	0.999	0.999	0.838	0.986	0.969	0.982	0.976	0.985
ca-HepTh	9877	31.36	45.2	75.32	65.17	90.88	93.71	0.999	0.999	0.984	0.996	0.99	0.996	0.99	0.996
ca-CondMat	23133	29.37	51.69	79.57	63.02	97.92	98.58	1	1	0.999	0.999	0.989	0.997	0.99	0.997
Enron Email	36692	35	48.88	73.54	70.42	98.88	99.45	1	1	0.999	0.998	0.983	0.99	0.98	0.99

Table 1. Estimated parameters of real-world networks using SLPA and LFME community detection algorithms to define the overlapping nodes. N is the size of the network. on is the fraction of overlapping nodes. S is the size of the neighborhood of the overlapping nodes. A_n is the proportion of common nodes of the two sets. A_t is the proportion of common nodes of the 10% of the most highly ranked nodes in the two sets. r represents the Rank-biased overlap between the set of neighbors of the overlapping nodes and the set of hubs. p determines the weights of the elements. More weight is given to the first elements of both sets if p has a small value. ρ is the Pearson correlation while ρ_s is the Spearman correlation.

Results show that there is a high overlap between the set of neighbors of the overlapping nodes and the set of hubs. It appears that a high proportion of the hubs are one-step neighbors of the overlapping nodes. One can notice from the values shown in Table 1 that the nodes belonging to the set of hubs with the highest degrees are always neighbors of the overlapping nodes (high values of A_t). Results show also that the Rank-biased Overlap between the set of neighbors of the overlapping nodes and the set of hubs has very high values. This confirms that there is a great similarity between the two sets. Moreover, there is a very high correlation between the set of neighbors of the overlapping nodes and the hubs. All these results are a strong indication that overlapping nodes are neighbors of the highly connected nodes of the network. Additionally, the influence of the neighborhood size of the overlapping nodes on the proportion of hubs has been investigated. Results show that increasing slightly the size of the neighborhood of the overlapping nodes allows to get a vast majority of hubs in this neighborhood. Furthermore, the degree distribution analysis of both sets shows that they display a similar empirical behavior. They exhibit a power-law degree distribution with very close exponent values. The comparison of the original network with its sub-network formed only by the overlapping nodes and the hubs (the so-called Overlap-Hub network) show that the average, median distance and diameter of the Overlap-Hub network are much smaller than the values of the original network. We also looked at the influence of the number of communities shared by the overlapping nodes (membership degree) on the proportion of hubs in their neighborhood. Results show that the proportion of hubs in the neighborhood of the overlapping nodes increases with the number of communities they belong to. All these findings confirm that the overlapping nodes are neighbors with a large proportion of hubs. These results are quite consistent, whatever the algorithm used to uncover the community structure. They may have implications in various contexts. They shed a new light on complex networks structuration that can be exploited for designing new immunization strategies (e.g., searching for hubs in large networks starting from overlapping nodes) and community detection algorithms.

¹ Ghalmane, Z., Cherifi, C., Cherifi, H., El Hassouni, M.: "Centrality in Complex Networks with Overlapping Community Structure", *Scientific Reports Journal* (2019), vol. 9, no 1. 2