The Clustering Methods in Graph Models of Social Networks

Tatiana Victorovna Zudilova
ITMO National Research University (ITMO University)
Software Development Department
197101 Saint Petersburg, 49 Kronverksky Pr., Russian Federation

Sergei Evgenievich Ivanov
ITMO National Research University (ITMO University)
Department of Intelligent Technologies in Humanities
197101 Saint Petersburg, 49 Kronverksky Pr., Russian Federation

Abstract

Analysis of social networks based on graph models allows determining the associated subgroups and visualizing the results. The basic methods of research communities in graph models of social networks are clustering methods: Modularity, Centrality, Hierarchical, Spectral, K-means, PAM (Partitioning Around Medoids). In the methods of searching for communities, the results essentially depend on the chosen metric - the measure of similarity of the participants. As a measure of similarity, we proposed to choose an averaged centrality, which takes into account a number of properties for participants. The new method for clustering graph models based on averaging of centrality is represented. As an example, clustering of a subgroup of a social network consisting of hundreds of participants is performed. The results of clustering for the proposed method are compared with the main methods. The similarities of the distinguished subgroups are verified by means of the cosine and correlation coefficients of vertex similarity, the Jaccard and Dice similarity coefficients. We proposed a new metric - the average coefficient based on the coefficients of vertex similarity. A comparative analysis of clustering results showed that the proposed method for clustering graph models has an accuracy in determining clusters that is comparable with methods - Spectral, Hierarchical, K-Means, and surpasses the accuracy of methods - PAM, Centrality, Modularity.
1 Introduction

The clustering of the social network allows identifying the structures and associations hidden in a social network and similar in features. As a result of clustering, participants in the social network are clustered on the basis of similar characteristics, and participants in different clusters are significantly different. Clustering social networks is necessary to understand the common features of participants in online interactions.

The social networks play an important role for people's opinions. For example, when conducting various electoral and promotional activities, it is necessary to conduct community analysis and allocate clusters. Analysis of social networks allows predicting the results of elections or advertising in the internet space.

The search of related communities in social networks, such as Facebook, LinkedIn, Twitter, VKontakte, etc. is an important task of network analysis [1-5]. Analysis of social networks based on graph models allows you to identify communities, find cohesive groups, analyze the connections of network participants and visualize the results [6-11].

A social network can be represented by a graph model, where the members of the network are the vertices of the graph, and the edges represent the connections between them.

Figure 1 shows the graph model of a subgroup in a social network, consisting of a hundred participants. In the graph model, each participant is represented by a vertex with a sequence number - a unique identifier of the network participant.

Fig.1. Graph model of a subgroup in a social network
Defining communities in social networks necessary for their classification, grouping of participants, hypothesis testing. The analysis of models can be produced by various methods of mathematical modeling [12-15]. The basic methods of research communities in social networks are clustering methods in graph models: Modularity, Centrality, Hierarchical, Spectral, K-means, PAM [16]. Clustering determines the degree of similarity between neighboring vertices in the cluster. For community search methods, the results essentially depend on the chosen metric - the measure of similarity or difference of participants. The different centralities as metrics are considered. The following centralities are known, centrality by: Degree, Closeness, Betweenness, Radiality, Eccentricity, PageRank, Status, Katz, and Eigenvector [17].

2 The methods for finding communities in graph models

We applied traditional methods of clustering in graph models: Modularity, Centrality, Hierarchical, and Spectral for the analysis of the graph model of a hundred participants, presented in figure 1. For the considered graph identifies three community-based method of Modularity. 

\[
\{\{2,3,4,5,9,10,12,18,19,21,26,28,29,30,36,38,44,45,49,52,55,56,57,58,60,62,64,65,66,74,76,78,82,85,86,88,91,93,95,96,97\}, \{8,11,14,15,16,17,23,24,25,27,33,35,47,51,53,54,61,63,69,71,73,75,77,81,84,89,90,94,99,100\}, \{1,6,7,13,20,22,31,32,34,37,39,40,41,42,43,44,46,48,50,59,60,65,67,68,70,72,79,80,83,87,92,98\}\}
\]

For considered graph the four communities based on the Centrality method were identified

\[
\{\{1,6,7,13,20,22,31,32,34,35,37,39,40,41,42,43,44,46,48,50,59,60,65,67,68,70,72,79,80,83,85,87,92,98\}, \{8,11,14,15,16,17,23,24,25,27,33,47,49,51,53,54,61,63,69,71,73,75,77,81,84,89,90,94,99,100\}, \{3,4,10,12,18,19,21,26,30,36,38,45,49,52,55,57,58,62,64,66,74,76,78,82,91,93,97\}, \{2,5,9,28,29,56,86,88,95,96\}\}
\]

The four communities based on the Hierarchical method were identified

\[
\{\{1,6,7,13,20,22,31,32,34,37,39,40,41,42,43,44,46,48,50,59,60,65,67,68,70,72,79,80,83,85,87,92,98\}, \{8,11,14,15,16,17,23,24,25,27,33,35,47,51,53,54,61,63,69,71,73,75,77,81,84,89,90,94,99,100\}, \{3,4,10,12,18,19,21,26,30,36,38,45,49,52,55,57,58,62,64,66,74,76,78,82,91,93,97\}, \{2,5,9,28,29,56,86,88,95,96\}\}
\]

For considered graph are defined six communities based on the Spectral method.

\[
\{\{3,4,10,12,18,19,21,26,30,36,38,45,49,52,55,58,62,64,66,74,76,78,85,91\}, \{8,11,15,17,23,25,27,33,51,53,61,63,69,71,75,81,84,90,94,99,100\}, \{3,4,10,12,18,19,21,26,30,36,38,45,49,52,55,57,58,62,64,66,74,76,78,82,91,93,97\}, \{2,5,9,28,29,56,86,88,95,96\}\}
\]

3 The methods for clustering graph models based on the analysis of centrality

For clustering methods, the search of cluster is performed based on a measure of similarity. The results of clustering essentially depend on the chosen
measures of similarity.

We consider methods for finding communities of graph models based on measures of centrality and apply them for clustering the graph model shown in figure 1.

The measure centrality degree (Degree Centrality) is defined as the number of connections vertices.

For social network, centrality by degree determines the importance of the participant, depending on the number of friends or the connections in the graph.

The centrality by degree for vertices in order on figure 1 is equal to:

\[ Cd = \{12, 13, 9, 12, 9, 18, 24, 22, 12, 20, 25, 20, 13, 16, 23, 18, 24, 18, 18, 22, 15, 25, 24, 19, 22, 22, 17, 11, 19, 21, 31, 26, 16, 31, 19, 30, 15, 24, 22, 32, 28, 30, 20, 15, 22, 16, 25, 22, 28, 31, 17, 22, 23, 17, 9, 13, 18, 17, 19, 33, 13, 24, 11, 15, 11, 21, 13, 24, 27, 24, 16, 9, 12, 25, 15, 11, 13, 15, 21, 24, 14, 19, 28, 16, 9, 22, 9, 11, 36, 22, 27, 11, 28, 11, 10, 14, 32, 24, 24\}

The five communities are identified on the basis of the list of centralities by the Hierarchical method of clustering:

\[ \{\{61\}, \{90\}, \{41, 98\}, \{32, 35, 37, 43, 51\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 36, 38, 39, 40, 42, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 91, 92, 93, 94, 95, 96, 97, 99, 100\}\]

The five communities based on the list of centralities are determined by the Optimize method:

\[ \{\{32, 35, 37, 41, 42, 43, 50, 51, 61, 70, 84, 90, 92, 94, 98\}, \{6, 10, 12, 16, 18, 19, 24, 26, 30, 31, 36, 44, 58, 60, 67, 80, 83\}, \{14, 21, 28, 34, 38, 45, 47, 52, 55, 59, 65, 72, 76, 79, 82, 85, 97\}, \{1, 2, 3, 4, 5, 9, 13, 29, 56, 57, 62, 64, 66, 68, 73, 74, 77, 78, 86, 88, 89, 93, 95, 96\}, \{7, 8, 11, 15, 17, 20, 22, 23, 25, 27, 33, 39, 40, 46, 48, 49, 53, 54, 63, 69, 71, 75, 81, 87, 91, 99, 100\}\]

The five communities are identified based on the list of centrality by K-Means method:

\[ \{\{14, 21, 34, 38, 45, 47, 65, 72, 76, 79, 82, 85, 97\}, \{32, 35, 37, 41, 42, 43, 50, 51, 61, 70, 84, 90, 92, 94, 98\}, \{6, 10, 12, 16, 18, 19, 24, 26, 30, 36, 44, 52, 55, 58, 59, 60, 83\}, \{1, 2, 3, 4, 5, 9, 13, 29, 56, 57, 62, 64, 66, 68, 73, 74, 77, 78, 86, 88, 89, 93, 95, 96\}, \{7, 8, 11, 15, 17, 20, 22, 23, 25, 27, 33, 39, 40, 46, 48, 49, 53, 54, 63, 67, 69, 71, 75, 80, 81, 87, 91, 99, 100\}\]

The five communities are determined based on the list of centrality by PAM method:

\[ \{\{14, 32, 35, 37, 41, 42, 43, 50, 51, 61, 69, 84, 90, 94, 98\}, \{6, 10, 12, 16, 18, 19, 24, 26, 30, 31, 36, 44, 58, 67, 80, 83\}, \{21, 28, 34, 38, 45, 47, 52, 55, 59, 60, 65, 72, 76, 79, 82, 85, 93, 97\}, \{1, 2, 3, 4, 5, 9, 13, 29, 56, 57, 62, 64, 66, 68, 73, 74, 77, 78, 86, 88, 89, 93, 95, 96\}, \{7, 8, 11, 15, 17, 20, 22, 23, 25, 27, 33, 39, 40, 46, 48, 49, 53, 54, 63, 70, 71, 75, 81, 87, 91, 92, 99, 100\}\]

The measure of centrality by closeness determines how quickly the information spreads from the participant of the network. The shortest path in the graph determines the distance between the two network members. Closeness centrality is calculated as the reciprocal of the remoteness of vertex.

Five communities are identified on the basis of the list of centralities by the hierarchical method:

\[ \{\{61\}, \{90\}, \{41, 98\}, \{32, 35, 37, 43, 51\}, \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 36, 38, 39, 40, 42, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 91, 92, 93, 94, 95, 96, 97, 99, 100\}\]
The five communities are defined on the basis of the list of centralities by the local optimization method:

\[
\{\{32,35,37,41,42,43,50,51,61,70,84,90,92,94,98\}, \{6,10,12,16,18,19,24,26,30,31,36,44,58,60,67,80,83\}, \{14,21,28,34,38,45,47,52,55,59,65,72,76,79,82,85,97\}, \{1,2,3,4,5,9,13,29,56,57,62,64,66,68,73,74,77,78,86,88,89,93,95,96\}, \{7,8,11,15,17,2,22,23,25,27,33,39,40,46,48,49,53,54,63,69,71,75,81,87,91,99,100\}\}
\]

The communities are identified based on the list of centralities by the K-means method:

\[
\{\{32,35,37,41,42,43,50,51,61,70,84,90,92,94,98\}, \{6,10,12,16,18,19,24,26,30,31,36,44,58,60,67,80,83\}, \{14,21,34,38,45,47,52,55,59,65,72,76,79,82,85,97\}, \{1,2,3,4,5,9,13,29,56,57,62,64,66,68,73,74,77,78,86,88,89,93,95,96\}, \{7,8,11,15,17,2,22,23,25,27,33,39,40,46,48,49,53,54,63,69,71,75,81,87,91,99,100\}\}
\]

Five communities are identified on the basis of the list of centralities by the PAM method:

\[
\{\{14,21,34,38,45,47,65,72,76,79,82,85,97\}, \{32,35,37,41,42,43,50,51,61,70,84,90,92,94,98\}, \{6,10,12,16,18,19,24,26,28,30,36,44,52,55,58,59,60,83\}, \{1,2,3,4,5,9,13,29,56,57,62,64,66,68,73,74,77,78,86,88,89,93,95,96\}, \{7,8,11,15,17,20,22,23,25,27,31,33,39,40,46,48,49,53,54,63,67,69,71,75,80,81,87,91,99,100\}\}
\]

The measure of centrality by Betweenness determines the importance of the network participant in the dissemination of information. Betweenness centrality is defined as the number of shortest paths between all pairs of participants that pass through the participant.

Five communities are identified on the basis of the list of centralities by the hierarchical method:

\[
\{\{54\}, \{60\}, \{90\}, \{10,28,44,55,65,80,85,91\}, \{1,2,3,4,5,6,7,8,9,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,45,46,47,48,49,50,51,52,53,55,56,57,58,59,61,62,63,64,66,67,68,69,70,71,72,73,74,75,76,77,78,79,81,82,83,84,86,87,88,89,92,93,94,95,96,97,98,99,100\}\}
\]

The five communities are defined on the basis of the list of centralities by the local optimization method:

\[
\{\{10,28,44,55,65,80,85,90,91\}, \{41,43,49,50,60,61,67,70,75,87,95\}, \{6,22,32,3,35,37,45,48,51,57,76,92,97,98\}, \{2,7,12,17,18,19,26,30,36,40,42,58,82,84,93,94\}, \{1,3,4,5,8,9,11,13,14,15,16,20,21,23,24,25,27,29,31,34,38,39,46,47,52,53,56,59,62,63,64,66,68,69,71,72,73,74,77,78,79,81,83,86,88,89,96,99,100\}\}
\]

The five communities are identified based on the list of centralities by the K-Means method:

\[
\{\{10,28,44,55,65,80,85,90,91\}, \{41,43,49,50,60,61,67,70,75,87,95\}, \{6,22,32,3,35,37,45,48,51,57,76,92,97,98\}, \{2,7,12,17,18,19,26,30,36,40,42,58,82,84,93,94\}, \{1,3,4,5,8,9,11,13,14,15,16,20,21,23,24,25,27,29,31,34,38,39,46,47,52,53,56,59,62,63,64,66,68,69,71,72,73,74,77,78,79,81,83,86,88,89,96,99,100\}\}
\]

The five communities are identified based on the list of centrality by the PAM method:

\[
\{\{41,43,49,61,67,70,75,87,95\}, \{10,28,54,55,60,65,80,85,90,91\}, \{6,12,17,18,22,3\}\}
\]
The measure of centrality by Radiality is determined by the distance to the vertices and the diameter of the neighborhood.

The five communities are defined on the basis of the list of centralities by the method of local optimization:

\{\{3, 5, 29, 56, 64, 73, 86, 89, 96\}, \{12, 18, 25, 28, 34, 36, 57, 79, 82, 83, 89, 93\}, \{6, 7, 8, 11, 15, 17, 23, 27, 30, 39, 45, 49, 53, 63, 65, 69, 71, 75, 81, 99, 100\}, \{1, 2, 4, 9, 13, 14, 16, 19, 21, 24, 26, 38, 47, 52, 58, 62, 66, 68, 74, 77, 78, 79, 89, 95, 97\}, \{10, 20, 22, 31, 32, 33, 35, 37, 40, 41, 42, 43, 44, 46, 48, 50, 51, 54, 55, 60, 61, 67, 70, 80, 84, 85, 87, 90, 91, 92, 94, 98\}\}

The measure of centrality in eccentricity is determined by the reciprocal of the maximum distance to the vertex.

The two communities are defined on the basis of the list of centralities by the method of local optimization:

\{\{1, 2, 3, 4, 5, 8, 9, 11, 13, 14, 15, 16, 17, 19, 21, 24, 25, 26, 27, 29, 34, 36, 38, 47, 52, 53, 56, 58, 6, 34, 66, 68, 69, 71, 72, 73, 74, 75, 77, 79, 81, 83, 86, 88, 89, 93, 95, 96, 97, 99, 100\}, \{6, 7, 10, 12, 18, 20, 22, 23, 28, 30, 31, 32, 33, 35, 37, 39, 40, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 54, 55, 57, 59, 60, 61, 62, 65, 70, 72, 73, 74, 75, 78, 80, 82, 84, 85, 87, 90, 91, 92, 94, 98\}\}

The centrality of the referential ranking (PageRank Centrality) is determined by counting the importance of the references to the vertex and depends on the fraction of participation of the remote vertices and the attenuation coefficient.

The five communities are defined on the basis of the list of centralities by the method of local optimization:

\{\{1, 3, 4, 13, 64, 66, 68, 73, 77, 79, 85, 89, 93\}, \{5, 6, 14, 27, 31, 34, 47, 53, 56, 57, 59, 60, 62, 67, 72, 74, 78, 83, 86, 88\}, \{2, 7, 9, 11, 18, 19, 22, 30, 33, 39, 48, 52, 63, 70, 71, 81, 84, 92, 94, 97, 99, 100\}, \{1, 2, 4, 9, 13, 14, 16, 19, 21, 24, 26, 27, 30, 39, 45, 49, 53, 63, 65, 69, 71, 75, 81, 99, 100\}, \{1, 2, 4, 9, 13, 14, 16, 19, 21, 24, 26, 27, 30, 39, 45, 49, 53, 63, 65, 69, 71, 75, 81, 99, 100\}\}

The Katz centrality measure is determined by the number of all vertices that can be connected.

The five communities are defined on the basis of the list of centralities by the method of local optimization:

\{\{6, 16, 24, 34, 44, 54, 59, 60, 72, 80, 83\}, \{32, 35, 37, 41, 42, 43, 50, 51, 61, 70, 74, 82, 84, 90, 92, 94, 98\}, \{1, 10, 12, 13, 14, 18, 19, 26, 30, 36, 47, 49, 52, 55, 58, 65, 68, 79, 85, 91\}, \{7, 8, 11, 15, 17, 20, 22, 23, 25, 27, 31, 33, 39, 40, 46, 48, 53, 63, 67, 69, 71, 75, 81, 87, 99, 100\}, \{2, 3, 4, 5, 9, 21, 28, 29, 38, 45, 56, 57, 62, 64, 66, 73, 74, 76, 77, 78, 82, 86, 88, 89, 93, 95, 96, 97\}\}

The measure of centrality by eigenvector is defined as the sum of the centrality of neighboring vertices divided by a constant - the largest eigenvalue.
The centrality of the eigenvector is greater for that participant who has more friends and is more central.

The five communities are defined on the basis of the list of centralities by the method of local optimization:

\{
{1,13,14,47,55,65,68,85},
{6,16,24,34,44,54,59,60,72,79,80},
{32,35,37,41,42,43,50,51,61,70,84,90,92,94,98},
{7,8,11,15,17,20,22,23,25,27,31,33,34,40,46,48,53,63,67,71,75,81,83,87,99,100},
{2,3,4,5,9,10,12,18,19,21,26,28,29,30,36,38,45,49,52,56,57,58,62,64,66,73,74,76,77,78,82,86,88,89,91,93,95,96,97}
\}

4 Clustering of graph models based on the method of averaging the centrality

The approach of clustering graph models based on the averaging method of centrality considered in the previous part is proposed. The averaged centrality is determined by the method of normalizing and averaging the centrality. In the social network, the averaged centrality makes it possible to take into account dozens of factors of network participants.

The averaged centrality for vertices in order on the graph (Fig. 1) is equal to:

\[ C_q = (0.439116, 0.190786, 0.380273, 0.246465, 0.193039, 0.538551, 0.613062, 0.538346, 0.197287, 0.455249, 0.568818, 0.545959, 0.469664, 0.489393, 0.53455, 0.511086, 0.580153, 0.367635, 0.335868, 0.595141, 0.461328, 0.682197, 0.5735, 0.517953, 0.515999, 0.370675, 0.521098, 0.262032, 0.204165, 0.565296, 0.581231, 0.707481, 0.618441, 0.588072, 0.595141, 0.496154, 0.684391, 0.580372, 0.694779, 0.672959, 0.457775, 0.522071, 0.730283, 0.67348, 0.198579, 0.571215, 0.332101, 0.531674, 0.660731, 0.695575, 0.523302, 0.558847, 0.374004, 0.437115, 0.231013, 0.683508, 0.45193, 0.556938, 0.73929, 0.558847, 0.495952, 0.428744, 0.221471, 0.626041, 0.317852, 0.474004, 0.523302, 0.476669, 0.643663, 0.558847, 0.259976, 0.522793, 0.627726, 0.659079, 0.198579, 0.601226, 0.198579, 0.474004, 0.84134, 0.583617, 0.707276, 0.519573, 0.629363, 0.278202, 0.21017, 0.206955, 0.721263, 0.558847, 0.558847)\]

The five communities are defined on the basis of the list of centralities by the method of local optimization:

\{
{1,3,10,13,18,19,21,26,52,58,64,65,68,73},
{2,4,5,9,28,29,56,66,74,76,82,86,88,95,96,97},
{22,32,35,37,40,41,42,43,48,50,51,54,55,60,61,67,70,80,85,90,92,98},
{6,8,12,14,15,16,24,25,27,34,36,38,47,53,59,62,72,77,78,79,83,89,93},
{7,11,17,20,23,30,31,33,39,44,45,46,49,57,63,69,71,75,81,84,87,91,94,99,100}
\}

Figure 2 shows the five communities defined on the basis of the list of centralities by the method of local optimization.
5 The verification of clustering for graph models

The large number of criteria for assessing the quality of clustering is used. We performed verification of clustering and verification of vertex similarity in communities determined on the basis of a list of averaged centrality by means of the cosine and the correlation coefficients, the Jaccard and the Dice coefficients. The cosine coefficient of vertex similarity is also known as the Salton coefficient. The cosine coefficient is defined as the number of common vertices neighbors divided by the geometric mean of their degrees. The Jaccard similarity coefficient is defined as the number of common neighbors of vertices divided by the number of vertices that are adjacent. The Dice similarity coefficient is defined as the number of common neighbors of vertices divided by half of the sum of the degrees of the vertices.

We analyzed the percentage of errors in determining the vertex belonging to the cluster by means of the cosine and correlation coefficients of vertex similarity, the Jaccard and Dice coefficients. The table shows the results of calculating the percentage of errors in clustering based on the coefficients of vertex similarity. On the basis of the coefficients of vertex similarity, the average coefficient of similarity is determined.

<table>
<thead>
<tr>
<th>Method of clustering</th>
<th>The error by the correlation coefficient</th>
<th>The error by cosine coefficient</th>
<th>The error by Jaccard coefficient</th>
<th>The error by Dice coefficient</th>
<th>The error by the average coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral</td>
<td>3</td>
<td>7</td>
<td>2.2</td>
<td>7.8</td>
<td>5.0</td>
</tr>
<tr>
<td>Averaged centrality</td>
<td>5.2</td>
<td>5.4</td>
<td>5.2</td>
<td>5.4</td>
<td>5.3</td>
</tr>
</tbody>
</table>
Clustering methods in graph models of social networks

Table 1. (Continued): Percentage of clustering errors

<table>
<thead>
<tr>
<th>Method</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Means</td>
<td>5.3</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>6.75</td>
</tr>
<tr>
<td>PAM</td>
<td>5.6</td>
</tr>
<tr>
<td>Centrality</td>
<td>7.8</td>
</tr>
<tr>
<td>Modularity</td>
<td>6.2</td>
</tr>
</tbody>
</table>

The proposed method of choosing as a metric of averaged centrality for clustering graph models has the accuracy in determining clusters that is comparable with methods: Spectral, Hierarchical, K-Means and surpasses the accuracy of methods: PAM, Centrality, Modularity.

From the table 1 it follows that the greatest accuracy for determining clusters is the methods: Spectral and Averaged centrality.

Conclusion

Methods of searching for communities in social networks are the methods of clustering in graph models: Modularity, Centrality, Hierarchical, Spectral, K-means, PAM. In the methods of finding communities, the results essentially depend on the chosen metric - the measure of similarity of the participants. We propose, as a measure of similarity, to select the averaged centrality, which takes into account the many properties of the participants. We presented a new approach for clustering graph models based on the averaged centrality. As an example, we performed the clustering of a subgroup of the social network, consisting of hundreds of participants, represented by a graph model. We compared the results of clustering for various methods and carried out a verification of the similarity measure of the distinguished subgroups by means of the cosine, correlation, the Jacquard, Dice coefficients of vertex similarity. We proposed a new metric - the average coefficient based on the coefficients of vertex similarity. A comparative analysis of clustering results showed that the proposed method for clustering graph models has an accuracy in determining clusters that is comparable with methods - Spectral, Hierarchical, K-Means, and surpasses the accuracy of methods - PAM, Centrality, Modularity.

References


Received: March 15, 2017; Published: April 12, 2017