International Journal of
Molecular Sciences

MDPI

Article

CytoCluster: A Cytoscape Plugin for Cluster Analysis and Visualization of Biological Networks

Min Li [1], Dongyan Li [2], Yu Tang [1], Fangxiang Wu [1,3] and Jianxin Wang [1,*]

[1] School of Information Science and Engineering, Central South University, Changsha 410083, China; limin@csu.edu.cn (M.L.); tangyu@csu.edu.cn (Y.T.); faw341@mail.usask.ca (F.X.W.)

[2] School of software, Central South University, Changsha 410083, China; dongyanli@csu.edu.cn

[3] Department of Mechanical Engineering and Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

[*] Correspondence: jxwang@mail.csu.edu.cn; Tel.: +86-731-888-30212

Received: 7 August 2017; Accepted: 23 August 2017; Published: 31 August 2017

Abstract: Nowadays, cluster analysis of biological networks has become one of the most important approaches to identifying functional modules as well as predicting protein complexes and network biomarkers. Furthermore, the visualization of clustering results is crucial to display the structure of biological networks. Here we present CytoCluster, a cytoscape plugin integrating six clustering algorithms, HC-PIN (Hierarchical Clustering algorithm in Protein Interaction Networks), OH-PIN (identifying Overlapping and Hierarchical modules in Protein Interaction Networks), IPCA (Identifying Protein Complex Algorithm), ClusterONE (Clustering with Overlapping Neighborhood Expansion), DCU (Detecting Complexes based on Uncertain graph model), IPC-MCE (Identifying Protein Complexes based on Maximal Complex Extension), and BinGO (the Biological networks Gene Ontology) function. Users can select different clustering algorithms according to their requirements. The main function of these six clustering algorithms is to detect protein complexes or functional modules. In addition, BinGO is used to determine which Gene Ontology (GO) categories are statistically overrepresented in a set of genes or a subgraph of a biological network. CytoCluster can be easily expanded, so that more clustering algorithms and functions can be added to this plugin. Since it was created in July 2013, CytoCluster has been downloaded more than 9700 times in the Cytoscape App store and has already been applied to the analysis of different biological networks. CytoCluster is available from http://apps.cytoscape.org/apps/cytocluster.

Keywords: biological networks; cluster analysis; cytoscape; visualization

1. Introduction

In recent years, people have paid more and more attention to recognizing life activities within a cell by protein interactions and protein complexes [1–3] in the field of systems biology. Proteins are one of the most important biological molecules in a cell. Within a cell, a protein cannot work alone, but rather works together with other proteins to perform cellular functions. Proteins are involved in a life process through protein complexes. Protein complexes can help us to understand certain biological processes and to predict the functions of proteins. Also, they can realize the cell signaling regulation functions by allosteric, competitive binding, interaction, and post-translational modification [4]. Protein-protein interaction (PPI) networks are powerful models that represent the pairwise protein interactions of organisms. Clustering PPI networks can be useful for isolating groups of interacting proteins that participate in the same biological processes or that, together, perform specific biological functions.

Up to now, many clustering algorithms, which are used to predict protein complexes from proteomics data, have been proposed and applied to biological networks. Out of these methods, the graph-based approaches are the most popular, which includes the partition-based clustering method, the density-based clustering method, the hierarchical-based clustering method and the spectral-based clustering method.

The partition-based clustering algorithms detect protein complexes by finding an optimal network partition, and making sure that the divided objects in the same cluster are as close as possible and the objects in different clusters are as far away as possible, such as HCS (Highly Connected Subgraph) [5], RNSC (Restricted Neighborhood Search Clustering) [6], MSCF (Minimal Seed Cover for Finding protein complexes) [7]. These partition-based clustering algorithms need to know the partition number, which is albeit generally unknown to us. What is more, partition-based methods cannot predict overlapping clusters.

The density-based clustering algorithms identify protein complexes by mining dense subgraphs from biological networks, such as MCL (Markov CLuster) [8], MCODE (Molecular COmplex DEtection) [9], CPM (Clique Percolation Method) [10], LCMA (Local Clique Merging Algorithm) [11], Dpclus (Density-periphery based clustering) [12], IPCA (Identifying Protein Complex Algorithm) [13], CMC (Clustering based on Maximal Cliques) [14], MCL-Caw (a refinement of MCL for detecting yeast complexes) [15], ClusterONE (Clustering with Overlapping Neighborhood Expansion) [16], and so on. These clustering algorithms have the advantage of recognizing dense subgraphs. However, it is difficult to predict the clusters which are non-dense subgraphs with these methods, such as the subgraph of "star" and "cycle."

The basic idea of the hierarchical clustering method is measuring the possibility that any two proteins are located in the same cluster according to their similarity or the distance between them. Hierarchical clustering methods can be further divided into divisive methods and agglomerative methods. A divisive method is a top-down approach, whose main action regards the total PPI network as a cluster first, then divides the network according to a rule until all nodes belong to different clusters. An agglomerative method is a bottom-up approach, whose main action regards each protein in the PPI network as a cluster first, then merges any two clusters according to their similarity value until all nodes are assigned to clusters. For example, G-N (Girvan-Newman) [17], MoNet (Modular organization of protein interaction Networks) [18], FAG-EC (Fast AGglomerate algorithm for mining functional modules based on the Edge Clustering coefficients) [19], EAGLE (agglomerativE hierarchicAl clusterinG based on maximaL cliquE) [20], HC-PIN (Hierarchical Clustering algorithm in Protein Interaction Networks) [21] are all hierarchical clustering algorithms. Hierarchical clustering methods can be used for mining arbitrary shape clusters, and can render the hierarchical organization of the entire PPI network based on a tree structure. However, this type of method is very sensitive to noise data and cannot obtain overlapping clusters. Some researchers extend the hierarchical clustering method to detect overlapping clusters by initializing a triangle with three interacting proteins instead of a single protein, such as OH-PIN (identifying Overlapping and Hierarchical modules in Protein Interaction Networks) [22].

The spectral-based clustering algorithms predict protein complexes based on the spectrum theory, such as QCUT (Combines spectral graph partitioning and a local search to optimize the modularity Q) [23], ADMSC (Adjustable Diffusion Matrix-based Spectral Clustering) [24], and SSCC (Semi-Supervised Consensus Clustering) [25]. These spectral-based clustering methods can be a simple and fast approach to a certain extent. These clustering algorithms depend on the feature vector, which determines the final clustering results. In addition, many other kinds clustering algorithms can be found in survey papers [26,27].

With the developments of clustering methods, the visualization of clusters becomes more and more important. Several tools [28–33] have been developed to help researchers to better recognize positive protein complexes. Cytoscape [34] is a friendly and open bioinformatics platform, which shows an exceptional performance both in virtualizations and manipulation of biological networks. Cytoscape also has the advantage of formidable extensibility of integrating a vast amount of plugins with diverse functions over other platforms. There are 33 apps concerning clustering based on Cytoscape described in our supplement, many of which aim to find meaningful pathways, or visualize networks by semantic similarities, or construct dynamic networks. Among all of the apps, there are several apps, such as ClusterViz [35], clusterMake [36], and ClusterONE [16], which are used to detect and visualize protein complexes in PPI networks. They are all useful tools with different clustering methods, which have been used in different areas of life sciences in recent years. However, a great deal of newly developed clustering algorithms has lost favor with the Cytoscape platform and do not implement visualization. Also, several plugins with old versions cannot work on the new Cytoscape platform any more. In order to solve the above limitations, we developed a new plugin named CytoCluster, which integrates six new clustering algorithms in total. In our plugin, five new approaches named IPCA, OH-PIN, HC-PIN, DCU (Detecting Complexes based on Uncertain graph model) [37], IPC-MCE (Identifying Protein Complexes based on Maximal Complex Extension) [38] were added, which are not integrated in any existing apps, but are important methods used to predict protein complexes. Our CytoCluster plugin also contains the BinGO function, which is used to determine which Gene Ontology (GO) categories are statistically overrepresented in a set of genes or a subgraph of a biological network. So, our app becomes a versatile tool that offers such comprehensive clustering algorithms, in addition to the BinGO function for biological networks.

2. Architecture

In this paper, we adopt Cytoscape 3.x to develop our app. Cytoscape 3.x has notable advantages over Cytoscape 2.x, which can be described in the following two aspects. First, the platform of Cytoscape 3.x adopts the OSGI (Open Service Gateway Initiative) framework, which allows developers to dynamically install, load, update, unload, and uninstall the newly developed bundles in an easy way. Second, Cytoscape 3.x employs Maven, which can help developers manage many jar files. In Cytoscape 3.x, both core modules and apps are called OSGI bundles, and they can significantly reduce complexity in app development to some extent. Also, two methods can be used for developing apps in Cytoscape 3.x. The first way is to develop apps as bundles, which can both register a service in the OSGI framework and withdraw its service from the registry. The second way is to implement the apps with Simplified CyApp API (Application Programming Interface), just like in Cytoscape 2.x.

The architecture of CytoCluster is shown in Figure 1, which includes three main bundles: the interface of CytoCluster bundles, the cluster algorithm bundles, and the visualization, BinGO, and export bundles. The interface of CytoCluster bundles is made up of a graphic user interface and a data exchange system, which allows the users to obtain different forms of bioinformatics networks including .txt and .csv files, and send the clustering results to Cytoscape. The six clustering algorithms bundles play an important role in our plugin CytoCluster, and we have defined the abstract Java class named clustering algorithms, making it is easy for us to integrate more clustering algorithms in CytoCluster. The BinGO bundles are the core functionality in analyzing the GO terms, which can be used to determine which GO categories are statistically overrepresented in a set of genes or a subgraph of a biological network. The visualization of BinGO and export bundles provide a way to intuitively visualize the clustering results in Cytoscape, determine which GO categories are statistically overrepresented, and export the clustering results to .txt or .cvs files.

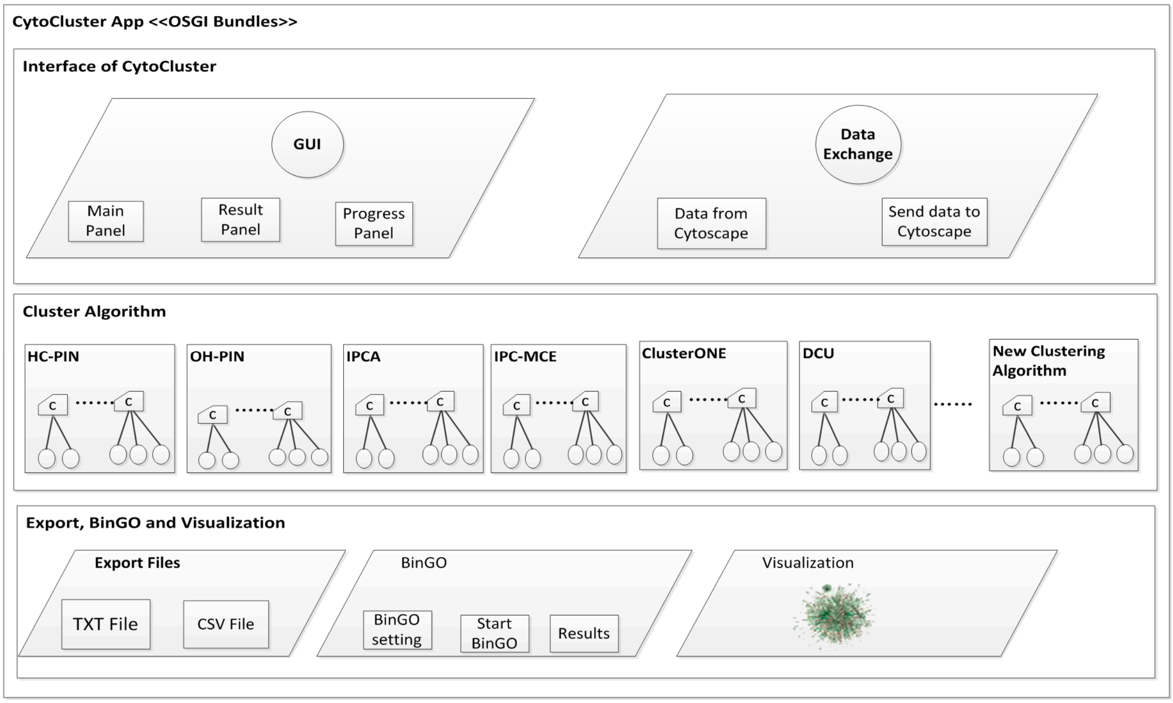

Figure 1. Architecture of CytoCluster.

3. Implementation

A user-friendly clustering software system to detect clusters is very important for biologists. By running the software, users can easily detect and analyze the protein complexes participating in the different life activities. Based on this basic idea, we developed our plugin CytoCluster by adopting the OSGI framework and the Cytoscape Maven archetypes. These frameworks and archetypes can create a maven-based project that builds an initial OSGI bundle-based Cytoscape app. The design is guided by the following three goals: first, to extend new clustering algorithms and add more functions; second, to dispatch the interface of CytoCluster and the algorithms; third, to respond quickly when the user operates the GUI (Graphical User Interface).

CyActivator class is an abstract class, which plays an important role in connecting Cytoscape with CytoCluster. All of the functions of CyActivator start to work as soon as you install the CytoCluster.jar for Cytoscape. The Analyze Action, as one of the service bundles, is the most important function in CytoCluster. Once the network is imported into Cytoscape, then our plugin CytoCluster is able to obtain these data from Cytoscape for further analysis. Two parts can be seen in the main panel. The top part mainly contains the two kinds of the clustering algorithms, overlap clustering algorithms and nonoverlap clustering algorithms. The bottom panel mainly provides six clustering algorithm panels, which are the IPCA panel, HC-PIN panel, OH-PIN panel, DCU panel, ClusterONE panel, and IPC-MCE panel. The user can choose different parameters according to their needs from these clustering algorithm panels. The result panel and the "export to .txt" function must be contained in CytoCluster, which provides an easy way to further analyze the results produced by different clustering algorithms. In addition, the progress panel is included in our app, which is used to visualize the progression of the running clustering algorithms.

Finally, we constructed this CytoCluster app containing four parts: Open, Close, About, and BinGO. Each part has its own function. Six clustering algorithms are included in the Open part. When users want to terminate this app, they should select the Close part. Here, BinGO plays an important role in determining which GO categories are statistically overrepresented in biological networks. Lastly, if you want to learn more information about the app, you cannot miss the About part.

3.1. Calculation and Basic Analysis

When users open the CytoCluster plugin, six clustering algorithms are provided, which are HC-PIN, OH-PIN, IPCA, IPC-MCE, ClusterONE, and DCU. In the following, these six clustering algorithms are briefly described.

3.1.1. HC-PIN (Hierarchical Clustering Algorithm in Protein Interaction Networks)

The HC-PIN algorithm [21] is a fast, hierarchical clustering algorithm, which can be used in a weighted graph or an unweighted graph. The main processes can be described as follows. First, all vertices in the PPI network are regarded as singleton clusters. Then, HC-PIN [21] calculates the clustering value of each edge and queues all of the edges into a queue Sq in non-increasing order according to their clustering values. The higher clustering value the edge has, the more likely its two vertices will be in the same module. In the process of adding edges in the queue Sq to cluster, λ-modules are formed. Finally, λ-modules can be outputted when the number of its proteins is no less than a threshold *s*.

3.1.2. OH-PIN (Identifying Overlapping and Hierarchical Modules in Protein Interaction Networks)

The OH-PIN algorithm [22] is an improved hierarchical clustering method, which can identify overlapping clusters. The basic idea of OH-PIN can be summarized as follows. At the beginning, the cluster set C_set is empty. For each edge in the protein interaction network, its B_Cluster is generated and the B_Cluster is added to the C_set, if B_Cluster is not already included in the C_set, until every B_Cluster is included. Then, OH-PIN [22] merges all highly overlapping cluster pairs in the C_set in terms of the threshold overlapping value. After the above step, OH-PIN assembles all of the clusters in the C_set into λ-modules by gradually merging the cluster pair with the maximum clustering coefficient.

3.1.3. IPCA (Identifying Protein Complex Algorithm)

The IPCA algorithm [13] is a density-based clustering algorithm, which can identify dense subgraphs in protein interaction networks. IPCA has four major sub-algorithms: weighting vertex, selecting weed, extending cluster, and extend-judgment. First, IPCA [13] calculates the weight of each edge by counting the common neighbors of its connected two nodes and computes the weight of each node by summing up the weights of its incident edges. The higher weight one node has, the more likely the node is regarded as the seed. At the beginning, a seed is initialed as a cluster. IPCA extends a cluster by adding vertices recursively from its neighbors in terms of the nodes' priority. Whether a node can be added to a cluster is determined by two conditions: its interaction probability and the shortest path between it and the nodes in the cluster.

3.1.4. IPC-MCE (Identifying Protein Complexes based on Maximal Complex Extension)

The IPC-MCE algorithm [38] is a maximal clique-based clustering algorithm. The basic idea of IPC-MCE can be described as follows. First, IPC-MCE removes all the nodes which have only one neighbor. Then IPC-MCE enumerates all the maximal cliques in the remained PPI network and puts them into the set MCS (Maximal Clique Sets). For each neighborhood vertex v of the maximal clique K in set MCS, if $\mathrm{IP}_{\mathrm{vk}}$ is no less than the threshold t, the vertex v can be added to the maximal clique K. The definition of $\mathrm{IPv}_{\mathrm{k}}$ is as follows:

$$\mathrm{IP}_{\mathrm{vk}} = \frac{|E_{\mathrm{vk}}|}{|V_k|} \tag{1}$$

E_{vK} is the number of the edges between the vertex v and K, and $|V_k|$ is the number of nodes in K. Finally, IPC-MCE [38] filters the repeated maximal clique according to a pre-defined overlapping value.

3.1.5. ClusterONE (Clustering with Overlapping Neighborhood Expansion)

The ClusterONE algorithm [16] mainly contains three steps. First, groups are grown by adding or removing vertices with high cohesiveness from selected seed proteins. At the beginning, the protein with the highest degree is regarded as the first seed and grows a cohesive group from it using a greedy procedure. ClusterONE repeats this grown process to form overlapping complexes until there are no proteins remaining in the PPI network. Then ClusterONE merges the highly overlapping pairs of locally optimal cohesive groups according to a pre-defined overlapping score. Finally, ClusterONE outputs protein complexes that contain no less than three proteins or whose density is larger than a given threshold ∂ (its default value is 0.8).

3.1.6. DCU (Detecting Complexes Based on Uncertain Graph Model)

The DCU algorithm [37] is a clustering algorithm, which detects protein complexes based on an uncertain graph model. First, DCU [37] starts from a seed vertex and adds other vertices by using a greedy procedure to form a candidate core with high cohesion and low coupling. Then, DCU uses a core-attachment strategy to add attachments to core sets to form complexes. Specifically, for each protein of a candidate set, if its internal absolute degree is less than its external absolute degree, which consists of neighbors of protein vertices in the candidate set, the protein must be removed from the candidate set. Finally, DCU needs to solve the problem of the repeated protein complexes by controlling their overlapping value. Users can select any kind of clustering algorithms they want in the main panel and input the parameters of the algorithm, which decide the creation of a specific clustering algorithm object in memory. Our CytoCluster plugin also provides the visualization of clustering results after running each of these six clustering algorithms, which can be seen in the result panel in the form of a thumbnail list. They can be sorted by the score, the size, or the modularity. In the result panel, the “Export” button and “Discard Result” button are included. The “Export” button is used for exporting results to a .txt file, including the name of algorithm, the parameters, and the clusters, while the “Discard Result” button is used for closing the result panel. Users can close the visualization of clustering results after running these six clustering algorithms with default parameters. In addition, users can see the visualization of cluster results after running a clustering algorithm. Therefore, CytoCluster is a convenient and fast app to obtain smaller networks from a large network.

3.2. BinGO

Here, we integrate the BinGO function to be the part of the CytoCluster. All this is done for the convenience of the users. When they install a cytocluster jar, users can not only choose different clustering algorithms, but also use BinGO. Once the BinGO part is opened, a panel will appear in the center of the computer monitor. Users can make a choice from this setting panel according to their need. The main function of BinGO is to determine the overrepresentation of Gene Ontology (GO) categories in a subgraph of a biological network or a set of genes. Once given a set of genes or a subgraph of a network on the GO hierarchy, BinGO can map the predominant functional themes and output this map in the form of a Cytoscape graph. The BinGO function has the same features as the BiNGO [39] plugin. These features contain graphs or genes list inputs; make and use custom annotations, ontologies, and reference sets; save the extensive results in a tab-delimited text file format; and so on. Selecting the “Start BiNGO” button is required after users have chosen their basic parameters. Then, the visualization of GO can be seen from a chosen network. The result can also be saved in a .bgo, which can be used for further studies.

In the BinGO part, two modes are included for selecting the set of genes to be functionally recommended. One is the default mode, and the other is the flexible mode. In the default mode, nodes can be chosen from a Cytoscape network, either manually or by other plugins. In the flexible mode, nodes can be selected from other sources, for example a set of nodes that are obtained from an experiment and pasted in a text input box. Here, the relevant GO annotations can be retrieved and propagated upwards through the GO hierarchy; namely, any genes related to a certain GO category can be predicted explicitly and included in all parental categories. Two statistical tests are also concerned so as to assess the enrichment of a GO term better. The most important characteristic of the BinGO part is its interactive use for molecular interaction networks, such as protein interaction networks. Furthermore, it is very flexible for BinGO to use ontologies and annotations. Both the traditional GO ontologies and the GOSlim ontologies are supported by BinGO. Then, the Cytoscape graph produced by BinGO can be seen, altered, and saved in a variety of ways.

4. Cases Studies

CytoCluster integrates different types of clustering algorithms including density-based clustering algorithms, hierarchical clustering algorithms, and maximal clique-based methods. Many researchers have downloaded and used the plugin since CytoCluster was released. So far, CytoCluster has been downloaded more than 9700 times since it was released in July 2013. Several important scientific articles indicated that CytoCluster can help scholars with their studies on the mechanisms of biological networks. There are several generic stages of how to run the clustering algorithms in our CytoCluster plugin, which include installing the CytoCluster app, loading the network, setting the data scope and parameters of clustering algorithms, running the cluster algorithm, and receiving or exporting the information of clustering results. The "CytoCluster" menu appears in the "App" menu, after installing the CytoCluster app. In this paper, we present a case to illustrate the use of our plugin. In addition, more cases on these six clustering algorithms can be seen in Table 1.

Table 1. More applications of CytoCluster and the six clustering algorithms integrated in it.

Algorithms	Application	Network	Description	Reference
IPCA	Exploring tomato gene functions	The tomato co-expression network was chosen and 465 complexes were found	IPCA was used to identity a densely connected network	[40]
	Unravelling gene function	The tomato co-expression network was chosen and 465 complexes were found	IPCA was choosen to identify thick connected nodes	[41]
	Predicting colon adenocarcinoma	The networks from IntAct and reactome were merged	IPCA was used to identify highly connected subnetworks	[42]
	The correlation between cold and heat patterns	The network from RA 18 was diagnosed with defciency pattern and 15 others were diagnosed with nondefciency pattern	IPCA was used to analyze the characteristics of networks	[43]
	Evidence-based complementary and alternative medicine	PPI network from genes was chosen so that the ratio of cold patterns to heat patterns in patients with RA was more or less than 1:1.4	IPCA was used to detect highly connected subnetworks	[44]
	Cold and heat patterns of rheumatoid arthritis	PPI network from these genes was chose that the ratio of cold patterns to heat patterns in patients with RA was more or less than 1:2	Highly connected regions associated with typical TCM cold patterns and heat patterns were identified	[45]
	Cold and heat pattern of rheumatoid arthritis	Network for differentially expressed genes between RA patients with TCM cold and heat patterns	IPCA was used to infer significant complexes or pathways in the PPI network	[46]
	Functional networks	Network contained some gene expressions or regulated proteins	Then eight highly connected regions were found by IPCA to infer complexes or pathways	[47]
	The molecular mechanism of interventions	PPI networks of biomedical combination was chosen and 11 complexes were found	IPCA was used to analyze the characteristics of the network	[48]
	The synergistic sechanisms	Network associated with Salvia miltiorrhiza and Panax notoginseng	Significant complexes or pathways were inferred	[49]
HC-PIN	Constraints on community	Associations between bacteria OTUs and four subnetworks were found	Subnetworks of OTUs were detected	[50]
	Strategies between two reef building cold-water coral species	Association network of the cold-water scleractinian corals bacterial communities	HC-PIN was used to identify OTUs	[51]
	Biomarkers	The network was extracted from the TCGA database	miRNA-gene clusters were identified	[52]
	Finding the candidate biomarkers for POAG disease	Network was extracted from previous studies with 474 proteins and nine subnetworks were found	HC-PIN was choosen to perform the clustering with a complex size threshold of 3	[53]
OH-PIN	Bacterial associations	Bulk soil DNA was extracted	The subnetworks were partitioned into modulars	[54]
ClusterONE	A census of human soluble protein complexes	Network was extracted from human HeLa S3 and HEK293 cells grown	ClusterONE was used to detect protein complexes	[55]
	An arabidopsis	A network with 8900 nodes and 6382 edges was chosen and 701 clusters were found	ClusterONE was used to obtain subnetworks	[56]
	Fndinge disease-drug modules	Disease-gene and drug-target associations were found from drug-target data	Overlapping subnetworks were identified	[57]

PPI: Protein-protein interaction; IPCA: Identifying Protein Complex Algorithm; TCM:Traditional Chinese Medicine; RA:Rheumatoid Arthritis; POAG: Primary Open Angle Glaucoma; OTU: Opearating Taxonomic Unit; TCGA:The Cancer Genome Atlas; OH-PIN: Identifying Overlapping and Hierarchical Modules in Protein Interaction Networks.

The case of CytoCluster was applied in botany [58]. This paper was published in Plant Physiology by Baute et al. The co-expression network was generated by Cytoscape 3.2.0 [59] according to the nodes and edges [60,61] at first. Then, the newly co-expression network was loaded, which incorporated 185 genes and 943 edges. Third, the main panel of the CytoCluster was opened and the HC-PIN clustering algorithm was chosen with standard settings and a complex size threshold of 10. In this case, 185 genes and 943 edges were included after dealing with the whole network. The identified subnetworks were further filtered, so as to only include the co-expression networks based on PCCs (Pearson Correlation Coefficients) of 0.7 and higher, as well as protein-protein interactions between query genes based on both experimental and predicted data from CORNET, when the users clicked on the "Analysis" button. Then, four subnetworks were formed after using our plugin for analysis, which can be seen from Figure 2. Each circle in Figure 2 shows a subnetwork. What is more, the generated co-expression network achieved by the HC-PIN algorithm can be seen in the result panel or exported to a .txt, so users can output the results from the different algorithms for further analysis. The table panel can list proprieties of clustering results when users select the corresponding clustering. The progress panel is used to visualize the progression of a specific cluster algorithm.

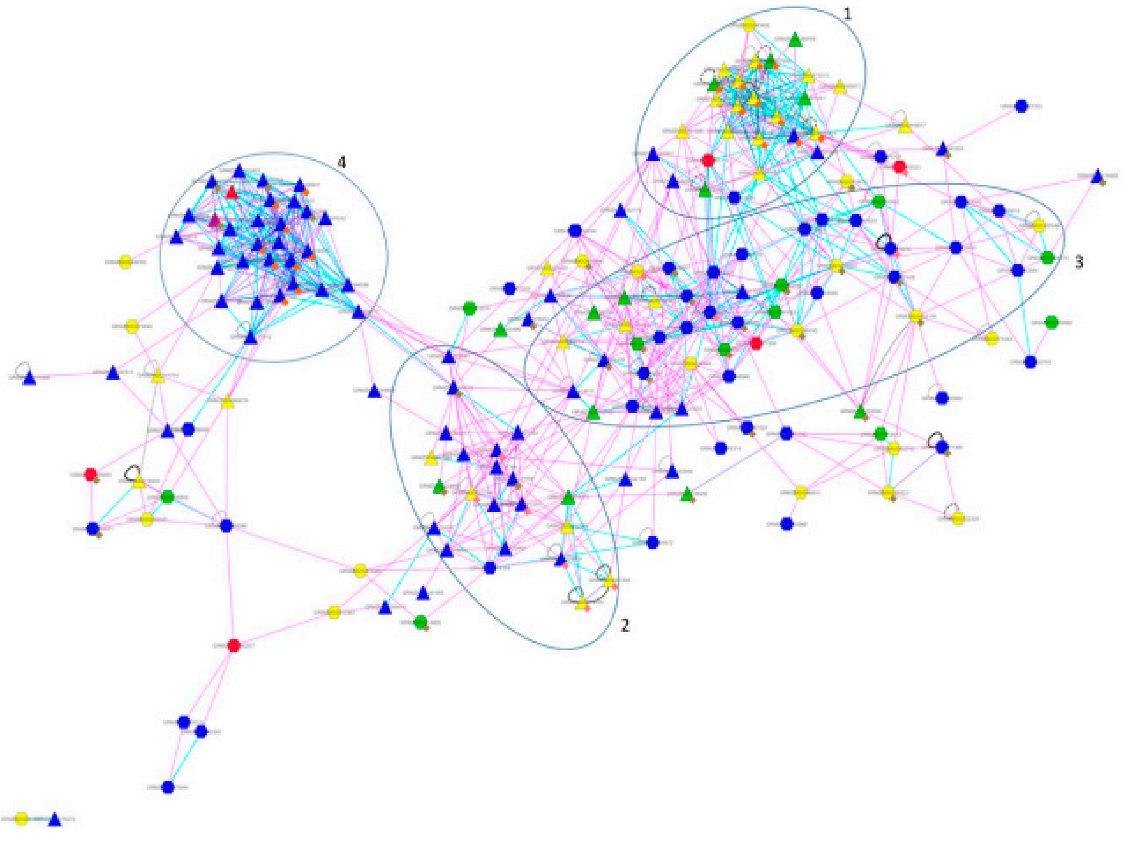

Figure 2. Four subnetworks achieved in the first case [58].

5. Conclusions

Our CytoCluster plugin is a platform-independent app for Cytoscape, which is also a functional diversity tool to offer different types of clustering algorithms, including IPCA, DCU, HC-PIN, OH-PIN, IPC-MCE, and ClusterONE. OH-PIN and HC-PIN are both hierarchical-based clustering algorithms, HC-PIN generates non-overlapping clusters, and on the contrary, OH-PIN produces overlapping clusters. IPCA, DCU, IPC-MCE, and ClusterONE are all density-based clustering algorithms, but the clusters generated by them also have some differences. Moreover, the same method will produce different results by changing the values of parameters. Users can both choose different clustering algorithms and analyze which GO categories are statistically overrepresented in a set of genes or a subgraph of a biological network. Our CytoCluster plugin is not only convenient for researchers to use, but also renders the investigated biological process easy to understand. Because our app has the advantage of expandability, more clustering algorithms such as those reported in References [62–65] as well as modules can be added to CytoCluster. Owing to such features, we firmly believe our app will be of great help in biology research.

Acknowledgments: This work was supported in part by the National Natural Science Foundation of China under Grants (No. 61622213, No. 61370024 and No. 61420106009).

Author Contributions: Min Li, Dongyan Li and Yu Tang conceived and designed the software, test and experiments; Dongyan Li and Yu Tang implemented the software and performed the experiments; Min Li and Dongyan Li wrote the paper. Min Li, Dongyan Li, Yu Tang, FangXiang Wu and Jianxin Wang revised the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, J.; Liang, J.; Zheng, W. A graph clustering method for detecting protein complexes. *J. Comput. Res. Dev.* **2015**, *52*, 1784–1793.
2. Alberts, B. The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell* **1998**, *92*, 291–294. [CrossRef]
3. Lasserre, J.P.; Beyne, E.; Pyndiah, S.; Pyndiah, S.; Lapaillerie, D.; Claverol, S.; Bonneu, M. A complexomic study of *Escherichia coli* using two-dimensional blue native/SDS polyacrylamide gel electrophoresis. *Electrophoresis* **2006**, *27*, 3306–3321. [CrossRef] [PubMed]
4. Gibson, T.J. Cell regulation: Determined to signal discrete cooperation. *Trends Biochem. Sci.* **2009**, *3410*, 471–482. [CrossRef] [PubMed]
5. Pržulj, N.; Wigle, D.A.; Jurisica, I. Functional topology in a network of protein interactions. *Bioinformatics* **2004**, *203*, 340–348. [CrossRef] [PubMed]
6. King, A.D.; Pržulj, N.; Jurisica, I. Protein complex prediction via cost-based clustering. *Bioinformatics* **2004**, *2017*, 3013–3020. [CrossRef] [PubMed]
7. Ding, X.; Wang, W.; Peng, X.; Wang, J. Mining protein complexes from PPI networks using the minimum vertex cut. *Tsinghua Sci. Technol.* **2012**, *176*, 674–681. [CrossRef]
8. Enright, A.J.; Dongen, S.V.; Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **2002**, *30*, 1575–1584. [CrossRef] [PubMed]
9. Bader, G.D.; Hogue, C.W.V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **2003**, *41*, 2.
10. Palla, G.; Derényi, I.; Farkas, I.; Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **2005**, *435*, 814–818. [CrossRef] [PubMed]
11. Li, X.L.; Foo, C.S.; Tan, S.H.; Ng, S.K. Interaction graph mining for protein complexes using local clique merging. *Genome Inform.* **2005**, *16*, 260–269. [PubMed]
12. Altaf-Ul-Amin, M.; Shinbo, Y.; Mihara, K.; Kurokawa, K.; Kanaya, S. Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinform.* **2006**, *7*, 207. [CrossRef] [PubMed]
13. Li, M.; Chen, J.; Wang, J.; Hu, B.; Chen, G. Modifying the DPClus algorithm for identifying protein complexes based on new topological structures. *BMC Bioinform.* **2008**, *9*, 398. [CrossRef] [PubMed]
14. Liu, G.; Wong, L.; Chua, H.N. Complex discovery from weighted PPI networks. *Bioinformatics* **2009**, *25*, 1891–1897. [CrossRef] [PubMed]
15. Srihari, S.; Ning, K.; Leong, H.W. MCL-CAw: A refinement of MCL for detecting yeast complexes from weighted PPI networks by incorporating core-attachment structure. *BMC Bioinform.* **2010**, *11*, 504. [CrossRef] [PubMed]
16. Nepusz, T.; Yu, H.; Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods* **2012**, *9*, 471–472. [CrossRef] [PubMed]

17. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [CrossRef] [PubMed]
18. Luo, F.; Yang, Y.; Chen, C.F.; Chang, R.; Zhou, J.; Scheuermann, R.H. Modular organization of protein interaction networks. *Bioinformatics* **2007**, *23*, 207–214. [CrossRef] [PubMed]
19. Li, M.; Wang, J.; Chen, J. A fast hierarchical clustering algorithm for functional modules in protein interaction networks. In Proceedings of the IEEE 2008 International Conference on BioMedical Engineering and Informatics (BMEI), Sanya, China, 27–30 May 2008; Volume 1, pp. 3–7.
20. Shen, H.; Cheng, X.; Cai, K.; Hu, M.B. Detect overlapping and hierarchical community structure in networks. *Phys. A Stat. Mech. Appl.* **2009**, *388*, 1706–1712. [CrossRef]
21. Wang, J.; Li, M.; Chen, J.; Pan, Y. A fast hierarchical clustering algorithm for functional modules discovery in protein interaction networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2011**, *8*, 607–620. [CrossRef] [PubMed]
22. Wang, J.; Ren, J.; Li, M.; Wu, F.X. Identification of hierarchical and overlapping functional modules in PPI networks. *IEEE Trans. Nanobiosci.* **2012**, *11*, 386–393. [CrossRef] [PubMed]
23. Chen, D.; Fu, Y.; Shang, M. A fast and efficient heuristic algorithm for detecting community structures in complex networks. *Phys. A Stat. Mech. Appl.* **2009**, *388*, 2741–2749. [CrossRef]
24. Inoue, K.; Li, W.; Kurata, H. Diffusion model based spectral clustering for protein-protein interaction networks. *PLoS ONE* **2010**, *5*, e12623. [CrossRef] [PubMed]
25. Wang, Y.; Pan, Y. Semi-supervised consensus clustering for gene expression data analysis. *BioData Min.* **2014**, *7*, 1–13. [CrossRef] [PubMed]
26. Li, M.; Wu, X.; Wang, J.; Pan, Y. Progress on graph-based clustering methods for the analysis of protein-protein interaction networks. *Comput. Eng. Sci.* **2012**, *34*, 124–136.
27. Ji, J.; Liu, Z.; Liu, H.; Liu, C. An overview of research on functional module detection for protein-protein interaction networks. *Acta Autom. Sin.* **2014**, *40*, 577–593.
28. Protein-Protein Interaction Networks Co-Clustering. Available online: http://wwwinfo.deis.unical.it/rombo/co-clustering/ (accessed on 21 April 2017).
29. Batagelj, V.; Mrvar, A. Pajek-program for large network analysis. *Connections* **1998**, *21*, 47–57.
30. Adamcsek, B.; Palla, G.; Farkas, I.J.; Derényi, I.; Vicsek, T. CFinder: Locating cliques and overlapping modules in biological networks. *Bioinformatics* **2006**, *22*, 1021–1023. [CrossRef] [PubMed]
31. Moschopoulos, C.N.; Pavlopoulos, G.A.; Schneider, R.; Likothanassis, S.D.; Kossida, S. GIBA: A clustering tool for detecting protein complexes. *BMC Bioinform.* **2009**, *10*, S11. [CrossRef] [PubMed]
32. Zheng, G.; Xu, Y.; Zhang, X.; Liu, Z.P.; Wang, Z.; Chen, L.; Zhu, X.G. CMIP: A software package capable of reconstructing genome-wide regulatory networks using gene expression data. *BMC Bioinform.* **2016**, *17*, 137. [CrossRef] [PubMed]
33. Li, M.; Tang, Y.; Wu, X.; Wang, J.; Wu, F.X.; Pan, Y. C-DEVA: Detection, evaluation, visualization and annotation of clusters from biological networks. *Biosystems* **2016**, *150*, 78–86. [CrossRef] [PubMed]
34. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]
35. Wang, J.; Zhong, J.; Chen, G.; Li, M.; Wu, F.X.; Pan, Y. ClusterViz: A Cytoscape APP for cluster analysis of biological network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2015**, *12*, 815–822. [CrossRef] [PubMed]
36. Morris, J.H.; Apeltsin, L.; Newman, A.M.; Baumbach, J.; Wittkop, T.; Su, G.; Bader, G.D.; Ferrin, T.E. clusterMaker: A multi-algorithm clustering plugin for Cytoscape. *BMC Bioinform.* **2011**, *12*, 436. [CrossRef] [PubMed]
37. Zhao, B.; Wang, J.; Li, M.; Wu, F.X.; Pan, Y. Detecting protein complexes based on uncertain graph model. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2014**, *11*, 486–497. [CrossRef] [PubMed]
38. Li, M.; Wang, J.X.; Liu, B.B.; Chen, J.E. An algorithm for identifying protein complexes based on maximal clique extension. *J. Cent. South Univ.* **2010**, *41*, 560–565.
39. Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **2005**, *21*, 3448–3449. [CrossRef] [PubMed]
40. Fukushima, A.; Nishizawa, T.; Hayakumo, M.; Hikosaka, S.; Saito, K.; Goto, E.; Kusano, M. Exploring tomato gene functions based on coexpression modules using graph clustering and differential coexpression approaches. *Plant Physiol.* **2012**, *158*, 1487–1502. [CrossRef] [PubMed]

41. Schaefer, R.J.; Michno, J.M.; Myers, C.L. Unraveling gene function in agricultural species using gene co-expression networks. *Biochim. Biophys. Acta (BBA)-Gene Regul. Mech.* **2017**, *1860*, 53–63. [CrossRef] [PubMed]
42. Wang, Y.; Zhang, J.; Li, L.; Xu, X.; Zhang, Y.; Teng, Z.; Wu, F. Identification of molecular targets for Predicting Colon Adenocarcinoma. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **2016**, *22*, 460–468. [CrossRef]
43. Wang, M.; Chen, G.; Lu, C.; Xiao, C.; Li, L.; Niu, X.; He, X.; Jiang, M.; Lu, A. Rheumatoid arthritis with deficiency pattern in traditional Chinese medicine shows correlation with cold and hot patterns in gene expression profiles. *Evid.-Based Complement. Altern. Med.* **2013**, *2013*, 248650. [CrossRef] [PubMed]
44. Lu, C.; Niu, X.; Xiao, C.; Chen, G.; Zha, Q.; Guo, H.; Jiang, M.; Lu, A. Network-based gene expression biomarkers for cold and heat patterns of rheumatoid arthritis in traditional Chinese medicine. *Evid.-Based Complement. Altern. Med.* **2012**, *2012*, 203043. [CrossRef] [PubMed]
45. Lu, C.; Xiao, C.; Chen, G.; Jiang, M.; Zha, Q.; Yan, X.; Kong, W.; Lu, A. Cold and heat pattern of rheumatoid arthritis in traditional Chinese medicine: Distinct molecular signatures indentified by microarray expression profiles in CD4-positive T cell. *Rheumatol. Int.* **2012**, *32*, 61–68. [CrossRef] [PubMed]
46. Chen, G.; Lu, C.; Zha, Q.; Xiao, C.; Xu, S.; Ju, D.; Zhou, Y.; Jia, W.; Lu, A. A network-based analysis of traditional Chinese medicine cold and hot patterns in rheumatoid arthritis. *Complement. Ther. Med.* **2012**, *20*, 23–30. [CrossRef] [PubMed]
47. Chen, G.; Liu, B.; Jiang, M.; Tan, Y.; Lu, A.P. Functional networks for Salvia miltiorrhiza and Panax notoginseng in combination explored with text mining and bioinformatical approach. *J. Med. Plants Res.* **2011**, *5*, 4030–4040.
48. Jiang, M.; Lu, C.; Chen, G.; Xiao, C.; Zha, Q.; Niu, X.; Chen, S.; Lu, A. Understanding the molecular mechanism of interventions in treating rheumatoid arthritis patients with corresponding traditional Chinese medicine patterns based on bioinformatics approach. *Evid.-Based Complement. Altern. Med.* **2012**, *2012*, 129452. [CrossRef] [PubMed]
49. Chen, G.; Liu, B.; Jiang, M.; Aiping, L. System Analysis of the Synergistic Mechanisms between Salvia Miltiorrhiza and Panax Notoginseng in Combination. *World Sci. Technol.* **2010**, *12*, 566–570.
50. Kalenitchenko, D.; Fagervold, S.K.; Pruski, A.M.; Vétion, G.; Yücel, M.; Le Bris, N.; Galand, P.E. Temporal and spatial constraints on community assembly during microbial colonization of wood in seawater. *ISME J.* **2015**, *9*, 2657–2670. [CrossRef] [PubMed]
51. Meistertzheim, A.L.; Lartaud, F.; Arnaud-Haond, S.; Kalenitchenko, D.; Bessalam, M.; Le Bris, N.; Galand, P.E. Patterns of bacteria-host associations suggest different ecological strategies between two reef building cold-water coral species. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **2016**, *114*, 12–22. [CrossRef]
52. Guo, H.; Chen, J.; Meng, F. Identification of novel diagnosis biomarkers for lung adenocarcinoma from the cancer genome atlas. *Orig. Artic.* **2016**, *9*, 7908–7918.
53. Atan, N.A.D.; Yekta, R.F.; Nejad, M.R.; Nikzamir, A. Pathway and network analysis in primary open angle glaucoma. *J. Paramed. Sci.* **2014**, *5*. [CrossRef]
54. Wang, H.; Wei, Z.; Mei, L.; Gu, J.; Yin, S.; Faust, K.; Raes, J.; Deng, Y.; Wang, Y.; Shen, Q.; Yin, S. Combined use of network inference tools identifies ecologically meaningful bacterial associations in a paddy soil. *Soil Biol. Biochem.* **2017**, *105*, 227–235. [CrossRef]
55. Havugimana, P.C.; Hart, G.T.; Nepusz, T.; Yang, H.; Turinsky, A.L.; Li, Z.; Wang, P.I.; Boutz, D.R.; Fong, V.; Phanse, S.; et al. A census of human soluble protein complexes. *Cell* **2012**, *150*, 1068. [CrossRef] [PubMed]
56. Van Landeghem, S.; de Bodt, S.; Drebert, Z.J.; Inzé, D.; van de Peer, Y. The potential of text mining in data integration and network biology for plant research: A case study on Arabidopsis. *Plant Cell* **2013**, *25*, 794–807. [CrossRef] [PubMed]
57. Wu, C.; Gudivada, R.C.; Aronow, B.J.; Jegga, A.G. Computational drug repositioning through heterogeneous network clustering. *BMC Syst. Biol.* **2013**, *7*, S6. [CrossRef] [PubMed]
58. Baute, J.; Herman, D.; Coppens, F.; de Block, J.; Slabbinck, B.; dell'Aqcua, M.; Pè, M.E.; Maere, S.; Nelissen, H.; Inzé, D. Combined large-scale phenotyping and transcriptomics in maize reveals a robust growth regulatory network. *Plant Physiol.* **2016**, *170*, 1848–1867. [CrossRef] [PubMed]
59. Czerwinska, U.; Calzone, L.; Barillot, E.; Zinovyev, A. DeDaL: Cytoscape 3 app for producing and morphing data-driven and structure-driven network layouts. *BMC Syst. Biol.* **2015**, *9*, 46. [CrossRef] [PubMed]

60. Kerrien, S.; Aranda, B.; Breuza, L.; Bridge, A.; Broackes-Carter, F.; Chen, C.; Duesbury, M.; Dumousseau, M.; Feuermann, M.; Hinz, U.; et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **2012**, *40*, D841–D846. [CrossRef] [PubMed]
61. Croft, D.; O'Kelly, G.; Wu, G.; Haw, R.; Gillespie, M.; Matthews, L.; Caudy, M.; Garapati, P.; Gopinath, G.; Jassal, B.; et al. Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res.* **2010**, *39*, D691–D697. [CrossRef] [PubMed]
62. Li, M.; Wang, J.; Chen, J.; Cai, Z.; Chen, G. Identifying the overlapping complexes in protein interaction networks. *Int. J. Data Min. Bioinform.* **2010**, *4*, 91–108. [CrossRef] [PubMed]
63. Li, X.; Wang, J.; Zhao, B.; Wu, F.X.; Pan, Y. Identification of protein complexes from multi-relationship protein interaction networks. *Hum. Genom.* **2016**, *10*, 17. [CrossRef] [PubMed]
64. Lei, X.; Ding, Y.; Wu, F.X. Detecting protein complexes from DPINs by density based clustering with Pigeon-Inspired Optimization Algorithm. *Sci. China Inf. Sci.* **2016**, *59*, 070103. [CrossRef]
65. Zhao, B.; Wang, J.; Li, M.; Li, X.; Li, Y.; Wu, F.X.; Pan, Y. A new method for predicting protein functions from dynamic weighted interactome networks. *IEEE Trans. Nanobiosci.* **2016**, *15*, 131–139. [CrossRef] [PubMed]