© 2018 IJSRSET | Volume 4 | Issue 9 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Comparative Network Pharmacology Based on Tanimoto Coefficient with Forbes-2 Coefficient

Basirun[1], Farit Mochammad Afendi[1,3], Wisnu Ananta Kusuma[2,3]

[1]Department of Statistics, Bogor Agricultural University, Bogor, Indonesia
[2]Department of Computer Science, Bogor Agricultural University, Bogor, Indonesia
[3]Tropical Biopharmaca Research Center Bogor Agriculture University, Bogor, Indonesia

ABSTRACT

Research on active compound contained in medicinal plant as ingredient for creating jamu has been widely practiced, but a detailed explanation of the mechanism of work in molecular and pharmacological still needs to be developed. In research of in silico, one of the common approaches done to look at the work mechanism of a compound was considering the similarity aspects of chemical structures between compounds. Measurement of similarity between compounds in general using Tanimoto coefficient. Based on the research result toward clustering 79 coefficient similarity to measure closseness of compounds, in addition Tanimoto, there was Forbes-2 coefficient found better similarity. Based on the statement the researcher was interested to do research with the aim of evaluating the Network Pharmacology medicinal plants that play the role of DM Type 2 by replacing Tanimoto coefficient with Forbes-2 coefficient. The evaluation method in this research used Mean Absolute Deviation (MAD). The result of the pharmacology network analysis using the Tanimoto coefficient was better compared to the Forbes-2 coefficient

Keywords : DM Type 2, Jamu, Forbes-2 Coefficient, Tanimoto Coefficient, MAD

I. INTRODUCTION

The development of rapid science had an impact on the paradigm shift in the healing of a disease from one drug target one which refers to a synthetic chemical drug containing an active chemical compound to treat a specific disease complaint into a "multi-component target network" involvement of several active compounds to target some of the proteins responsible for disease Zhang *et al* [1]. This new paradigm spawned several studies, especially in the search for active compounds in medicinal plants for the formulation of herbs. Computational statistical approaches can be used to obtain compounds that play an active role in treating a particular disease based on its target protein. Result of research conducted by Nurishmaya [2] and based on herb concoction which is being developed at Tropica Biopharmaca Research Centre Bogor Agriculture University, found there are 4 plants that acts as an effort in curing Diabetes Mellitus (DM) type 2, the plant are Bitter melon (*Momordica charantia*), Sembung (*Blumea balsamifera*), Bratawali (*Tinospora crispa*), and Ginger (Z*ingiber officinale*). Jamu herbs were chosen as an effort to uncover the mechanism of the work of *jamu* herbs that were proven efficacy as the *jamu* herbs of antidiabetes through testing in the laboratory. The efficacy was proven by the model of PLSDA, support vector machine (SVM) method and coefficient PLSDA multiways and used zebra fish as the a test animal.

Based on the results of Nurishmaya [2], Qomariasih [3] conducted a data analysis using pharmacological networking method called drugCIPHER and simultaneous clusrering. The drugCIPHER method was

performed to find target proteins that have the greatest correlation with each compound. The first step was to find the value of proximity between 74 compounds using Tanimoto coefficient. On the other hand, graphical analysis is performed to obtain the closest distance between proteins which are then converted to proximity between proteins through the concept of exponential calculations. Then the second step is to find the correlation between the two proximity values to produce 100 target proteins with the highest correlation with each of the 74 compounds named as DrugCIPHER's concordant scores. Each compound had high correlation with the same protein, in order to be taken only the unique protein. There were 1250 unique proteins that originally amounted to 7400 protein-protein connections. A concordant score of 74 compounds with 1250 proteins was then used in a simultaneous clustering to see the resemblance of a target protein profile with a compound. The results of simultaneous clustering showed that there were 13 potentially antidiabetic drugs, 2 of the Bratawali compounds and 11 of the Ginger plant compounds.

On the other hand Bakri [4] conducts clustering 79 coefficients of binary data resemblance. The coefficients of the binary data similarities are grouped to see the similarity between coefficients. The method of welding used was the Ward method. The result of clustering evaluated using ROC by looking at the value of AUC. The coefficient of binary data resemblance which results in a high AUC value is the coefficient of similarity of Forbes-2 binary data. Based on these findings, this study aims to evaluate the Network Pharmacology of Jamu that play a role in treating DM Type 2 by replacing the Tanimoto coefficient with the coefficient of Forbes-2.

II. RESEARCH METHOD

A. Data

The data used is the list of active material of medicinal herbs ingredients that has been being developed in the Tropica Biopharmaca Research Centre Bogor agriculture University for treatment of Diabetes Mellitus type 2 that consist of Bitter melon (*Momordica charantia*), Sembung (*Blumea balsamifera*), Bratawali (*Tinospora crispa*), and Ginger (*Zingiber officinale*). The data used by the researcher was the data that has been investigated by Qomariasih [3], the more detail list that related to resource of data like shown in Table 1. Based on Table 1 the number total of compound that would be analysed as much as 74 compounds namely 55 compounds of plants and 19 compounds of synthetic.

Table 1. The list of the data pangkalan

Data	Pangkalan Data	The Result of Searching Data
Compound	Take Out "Jamu" of Knapsack	Obtained 595 plant compounds consisting of 291 Ginger compounds, 18 Bratawali compounds, 41 compounds Sembung, and 245 Bitter melon compounds
	DrugBank Database	There were 19 active compounds of synthetic antidiabetic drugs
Active compound	Pubchem	Of the total of 595 compounds there are 58 active compounds that have biological activity as well as more detailed information such as SMILES and PubChem ID (CID) but which have intraction proteins as much as 55 active compounds
Fingerprint Klokata Roth (KR)	ChemDes	There are as many as 4860 substructures of molecular compounds
Interaction between protein	Human Protein Reference Database (HPRD)	There are 39240 binary interactions between 9673 unique proteins in the body
Conversion of Gi Number	Biological DataBase network	There are 462 unique proteins targeted from the active ingredient and 43 unique proteins that are targeted from synthetic compounds so that the total amount is 505 target proteins

Information 1 : Compound has criteria KR, 0: compound has no criteria of Kletkota Roth (KR) [5]

A. Method of Analysis

Target Protein Prediction

This stage uses the drugCIPHER method with the following steps[6]

1. Obtaining the matrix of drugCIPHER-CS

a. Creating a 4860 bit binary set for 74 compounds based on fingerprintKR

Table 2 Set of binary of two compounds

	Bit 1	Bit 2	..	..	Bit 4860
Compound 1	1	0	..	..	0
Compound 2	0	1	..	..	1
..	..	..	..	..	..
Compound 74	1	0	..	..	1

b. Calculating the vealue of similarity of structure of molecule by using Forbes-2 coefficient

$$CS_{Forbes-2} = \frac{\text{na} - (\text{a} + \text{b})(\text{a} + \text{c})}{\text{n}\min(a+b, a+c) - (a+b)(a+c)}$$

Information:

a = the value of both items is 1

b,c = item value is 1 and 0

d = value of both items is 0

n = total value

The value of the similarity coefficient produced in the form of a matrix of similarity of the compound with the value (0<*Forbes-2*<1).

2. Calculating drugCIPHER-GR (φ_{pd})

a. Calculates the closest distance between targeted protein (462 unique proteins from plant compounds and 43 drug proteins) with all the proteins in the body (9673 proteins) documented in HPRD databases using graph theory. So the shortest distance matrix formed is 505 × 9673.

b. Calculating drugCIPHER-GR

$$\varphi_{pd} = \sum_{pk \epsilon T(d)} e^{-L^2 ppk}$$

φ_{pd} is similarity between protein p and compound d on the basis of PPI, pk is the intended target protein by compound d, L^2ppk is the shortst distance between p and pk on the network of PPI in the genomic space, $e^{-L^2 ppk}$ is the used to convert the distance of proteins into proximity of proteins. From these calculations will be produced proximity matrix compounds with proteins with size 74×9673.

c. Calculating the value of the concordant score is the correlation of the proximity between the compound and the protein in the human body

$$\rho_{pd}^{C} = \frac{cov(CS_d, \varphi_{pd})}{\sigma\,(CS_d)\sigma(\varphi_{pd})}$$

From the above calculation will produce a matrix of 74 compounds × 9673 protein. For each compound will be taken 100 largest protein with the highest concordant value. Of the 100 highest protein will get as many as 967 unique target proteins that will form a matrix 74 × 967.

B. Pharmacology Network Evaluation

From the stages of estimating target proteins using the Pharmacology Network (drugCIPHER) to be evaluated is at stages 1 and 3 that is seeing the similarity between the compounds and the proximity of compounds with proteins in the body if the *coefficient of Tanimoto* is replaced by the coefficient of forbes-2 as Figure 1 As to what will be observed is how much deviation between the corresponding elements of the calculation Tanimoto coefficient and forbes-2 coefficient by using *Mean Absolute Deviation* (MAD).

$$MAD = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}\left|x_{ij} - y_{ij}\right|$$

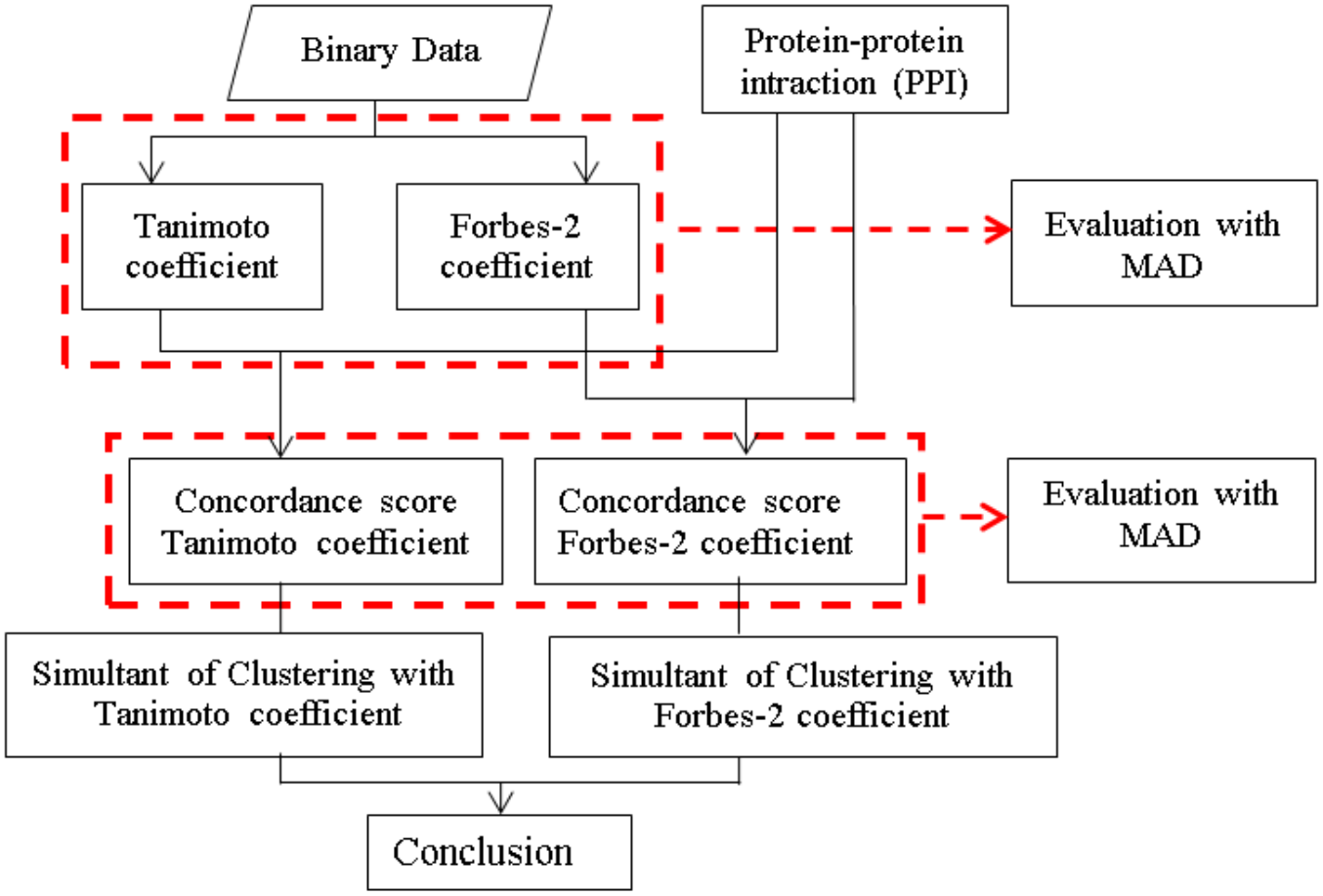


Figure 1. Evaluation steps of prediction of targetted protein

Information:

x_{ij} = coefficient value of similarity and concordance score on line (i) and column to (j) is calculated using Tanimoto coefficient

y_{ij} = coefficient value of similarity and concordance score on row (i) and column to (j) is calculated using Forbes-2 coefficient

m = number of rows

n = number of columns

C. Two Dimensional Simultant of Clustering

As many as 100 of the highest target proteins of 74 compounds that have been obtained will be selected concordant values of the unique protein alone, which is as many as 967 proteins. Then from the concordance value with size 74 × 967 done clustering 2 dimensional using and using method Ward.D1.

III. RESULT AND DISCUSSION

A. The Pharmacology Network

Jamu is one of the alternative herbs that can be used to treat diabetes type 2. The four plants that are predicted as the ingredients of making herbs are Bratawali, Ginger, Sembung, and Bitter melon, each containing compounds that play an active role in targeting certain proteins in the body. The four plants have 58 compounds, 868 target proteins and 416 unique proteins. Each compound can target several protents so that the number of connections between compounds with protein as much as 3059. Next connection between compounds and proteins form a network called the pharmacological network as in Figure 2.

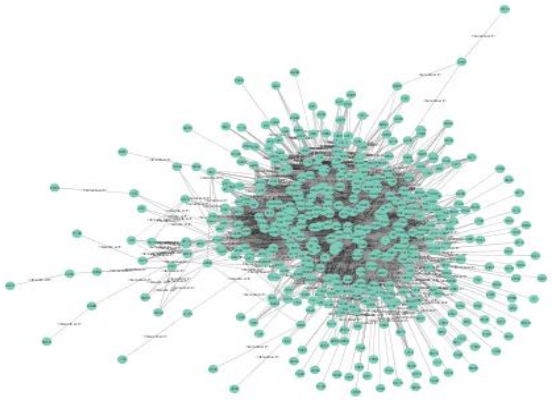
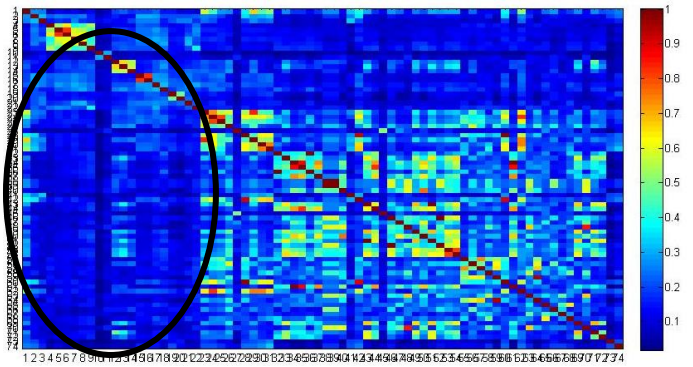


Figure 2. Compound network with protein

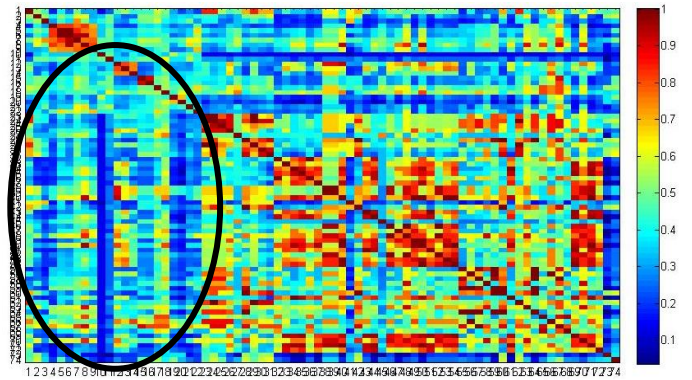
Based on Figure 2 the number of connection of compound of Bratawali plant as much as 70 protein, the connection of compound of Bitter melon plant as much as 47 protein, the connection of compound of Ginger plant as much as 2394 protein and the connection of compound of Sembung plant as much as 548 protein.

B. Targeted Protein Prediction

Determination of active antidiabetes melitus compounds in this study using the method of drugCIPHER. This method can turn the network into a numerical value. In addition, this method is very active role in determining the similarity between the compound either the plant compounds or synthetic compounds. The results of the search in the database showed that the number of compounds that have a connection with protein intraction protein (PPI) as many as 55 of 58 plant compounds and three plant compounds have no connection with PPI (J197, P185, S001). The number of compounds to be analyzed as many as 55 compounds from medicinal plants that are connected with PPI and 19 compounds derived from synthetic antidiabetic compounds that have been validated by the drug and food bodies of the United States amounting to 74 compounds. Target proteins can be seen from similarities between compounds. According to Johnson and Maggiora [7], compounds with similar chemical structures have similar biological properties. The similarity interval of the resulting value ranges from 0 to 1 is the greater the value of similarity then the compound will bind the protein more strongly and vice versa. In this study the compounds with low proximity are indicated by a dark blue color while the compound with high proximity is indicated by dark red. Visualize the result as shown in Figure 3.



(a)

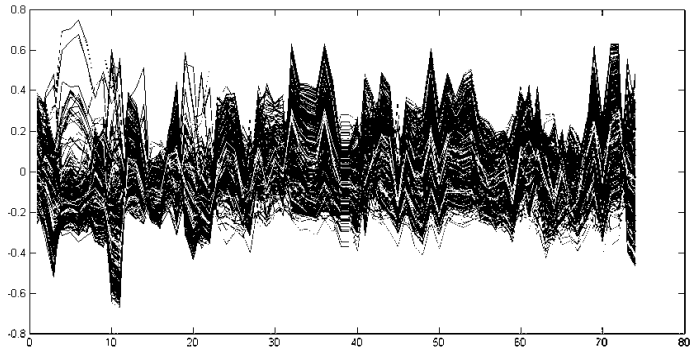


(b)

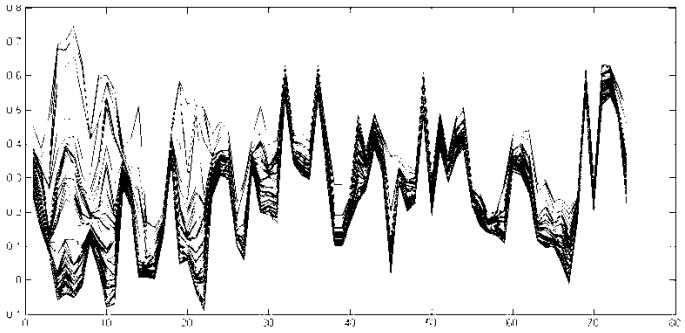
Figure 3. (a) Measurement of Similarities between compounds using Tanimoto coefficients and (b) measurement of similarities between compounds using Forbes-2 coefficients

The result of evaluation using *Mean Absolute Deviation* (MAD) shows that big deviation of similarity between compound with compound of 0.2586 means proximity of compound with compound as measured by using coefficient of Forbes-2 gives higher result. The proximity of the compound with the compound using Forbes 2 coefficient is more dominantly red and close to 1 as shown in the compound of DB02 with DB03, J010 with J036 and J127. In addition to the black of a circle in Figure 3 it is seen that with the coefficient of Forbes-2 there is a high value similarity between the plant compounds and the synthetic compounds that do not occur in the calculations using the Tanimoto coefficient.

The second step is to calculate the correlation value (concordant score) in the Pharmacology chamber and the genomic chamber. The value of a concordant score ranges from -1 to 1. The more positive the concordance score then each compound tends to target more and more specific proteins. Each compound has a concordant score of 9673 proteins in the body but it will take 100 proteins with the highest concordant value of each compound for further analysis. The collection of 100 highest protein values is assumed to be able to represent the profile of all proteins in the body such as Figure 4.



(a)



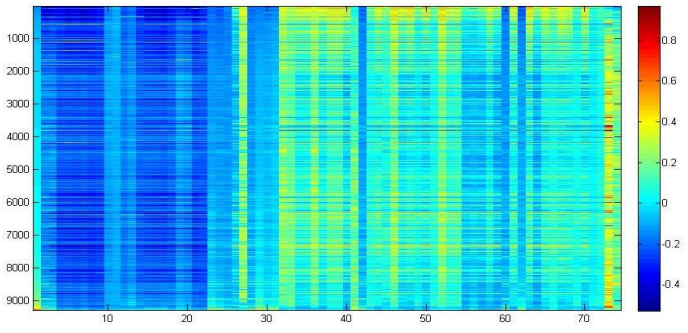
(b)

Figure 4 (a) a concordance score between 74 compounds with 9673 proteins in the body and (b) a concordant score of 74 compounds with 100 of the highest protein

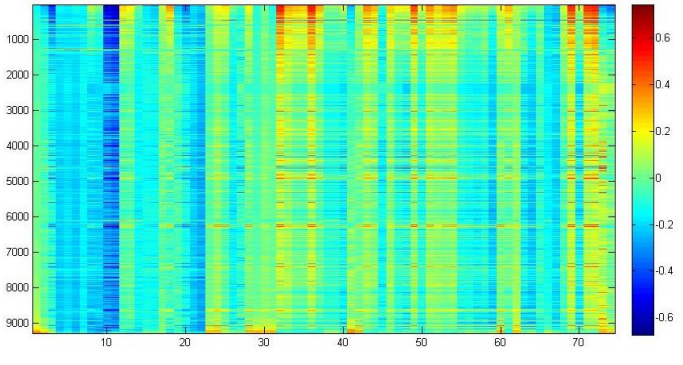
After searching the concordant score, the result of evaluation with MAD is smaller that is the big deviation of proximity between 74 compound with 9673 protein equal to 0.1392 and big deviation proximity between 74 compound with 100 highest protein equal to 0.1550 visualization see Figure 5. MAD evaluation result for concordant score show the ability of active compounds to target target proteins tends to be less visible than the number of unique target proteins produced by fewer than 1250 unique target proteins obtained with Tanimoto coefficient being 967 unique target proteins with Forbes 2 coefficient. Among the 967 unique target proteins are 19 out of 26 a protein that causes diabetes mellitus type 2 that has been targeted by 19 active compounds as an antidiabetic drug.

C. Two Dimensional Simultant of Clustering

The 2-dimensional hierarchy of bundles in this study used the Ward.D1 clustering method to produce different compounds and proteins. as shown in Figure 6. Dendograms of the lateral side are the result of clustering of proteins, while horizontal side dendogram shows the result of clustering of compounds. Bright colors indicate the correlation between sub-row matrix and sub column matrix has a high value whereas a rather dark color indicates that the correlation between sub row matrix and sub column matrix has a low value. Therefore, based on the suitability of color, the left compound girder more tends to correspond to the activity of the protein in the 4th quadrant whereas the right compound girder is more compatible with the protein activity in the 2nd quadrant.



(a)



(b)

Figure 5. Comparison between (a) score of concordant of compound with protein calculated by using Tanimoto coefficient and (b) score of concordant of compound with protein calculated by using Forbes-2 coefficient

The Second quadrant found 21 number of compounds were clustered, three of them were syntetic compound, 15 Ginger compound, 2 bitter melon compound and 1 sembung compound. In addition, it was also found as much as 490 protein and 4 of them are the cause of diabetes. The 3 synthetic compound was because of having the same role namely as hindrance enzin glucosidase.

The hindrance of enzyme can reduce the speed of carbohydrate disgestion complexly and reduce the absorbtion of glucose because the carbohydrate not broken in glucose molecule. To the patient who suffer diabetes, the effect of short term from therapy of this medicine was to reduce the degree of blood glucose, while the effect of long term was to reduce the degree of hemoglobine. Quadrant 4 found 53 number of compound was clustering, 16 of them was synthetic compound, 31 Ginger compound, 1 Bitter melon compound, Sembung compound and 3 Bratawali compound. While B018 clusters with some synthetic compound, also with some Ginger compound namely J288, J287, J271, and J068 cluster with some synthetic compound.

The analysis of simultant clustering give subset submatric which state that activity of protein only affected by some compound. The box of vertical showed that compound B108, compound DB17 and DB13 involved in some activities of protein namely 477 protein (characterized with the clear color) and 15 of them was the cause of Diabetes Melitus type 2.

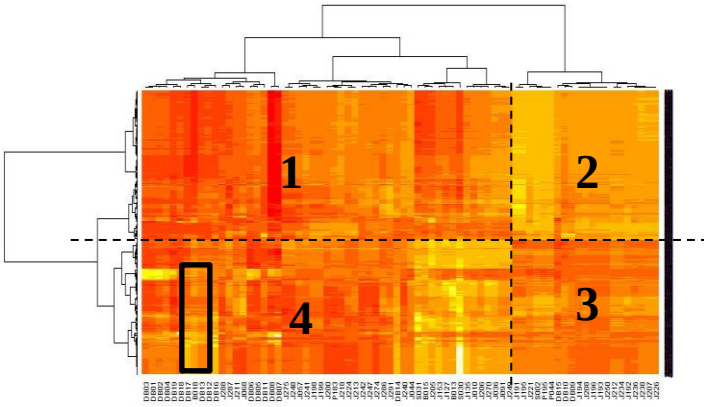


Figure 6. Clustering of simultant with method of Ward.D1

IV. CONCLUSION

Based on the result and discussion can be concluded some things about comparison of Pharmacology network based Tanimoto coefficient to Forbes-2 coefficient namely as follows:

1. The similarity between compounds that had been calculated by using Forbes-2 coefficient was higher.
2. After being done calculation score of concordant, the ability of active compound target unique target protein with Forbes-2 coefficient was getting smaller namely from 1250 unique target protein to be 967 unique target protein.
3. With Forbes-2 coefficient, the number of indicated protein to heal diabetes melitus type 2 smaller namely from 22 protein to be 19 protein.
4. With Tanimoto coefficient, it was found two compounds of plant that joining clustered in a quadrant with 19 synthetical compound while with Forbes-2 coefficient, the synthetical compound was separated into two quadrant in order to Tanimoto coefficient better than Forbes-2 coefficient.

V. REFERENCES

[1]. Zhang Gui-Biao, Li Qing-Ya, Qi-long Chen, Shi-bing Su. 2013. Jejaring Pharmacology: a new approach for chinese Jamu research. *Hindawi*. Article ID 621423, 1-9.

[2]. Nurishmaya MR. 2014. Pendekatan Bioinformatika Formulasi Jamu Baru Berkhasiat Antidiabetes dengan Ikan Zebra *(danio rerio)* sebagai Hewan Model [Skripsi]. Bogor(ID): Institut Pertanian Bogor.

[3]. Qomariasih N, Susetyo B, Afendi FM. 2016. Analisis gerombol simultan dan jejaring farmakologi antara senyawa dengan protein target pada penentuan senyawa aktif jamu anti Diabetes Tipe 2. Jurnal Jamu Indonesia. 1(2) : 30-40.

[4]. Bakri R, Wijayanto H, Afendi FM. 2016. Prediksi senyawa aktif pada tanaman obat berdasarkan kemiripan struktur kemiripan kimiawi untuk penyakit Diabetes Tipe 2. Jurnal Jamu Indonesia. 1(3) : 1-5.

[5]. Klekota J, Roth FP. 2008. Chemical Substructures that Enrich for Biological Activity. Bioinformatics, 24:2518-2525.

[6]. Zhao S, Li S. 2010. Network-based relating pharmacological and genomic spaces for drug target identification. *PLoS ONE*. 5(7): e11764. doi:10.1371/journal.pone.0011764.

[7]. Johnson AM, Maggiora GM. 1990. Concepts and Applications of Molecular Similarity. New York: John Willey&Sons.ISBN 0-471-62175-7.