



(51) International Patent Classification:

G16H 50/70 (2018.01) G16H 10/40 (2018.01)

(21) International Application Number:

PCT/US2019/031475

(22) International Filing Date:

09 May 2019 (09.05.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

15/983,235 18 May 2018 (18.05.2018) US

(71) Applicant: **BECKMAN COULTER, INC.** [US/US]; 250 South Kraemer Boulevard, Brea, CA 92821 (US).(72) Inventors: **MENHARDT, Wido**; 17706 Blackberry Hill Road, Los Gatos, CA 95030 (US). **DAVIS, Martha, C.**; 1082 Lynwood Drive, Brea, CA 92821 (US).(74) Agent: **MORRISS, William, S.** et al.; Frost Brown Todd LLC, 3300 Great American Tower, 301 East Fourth Street, Cincinnati, OH 45202 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: TEST PANEL ANALYSIS

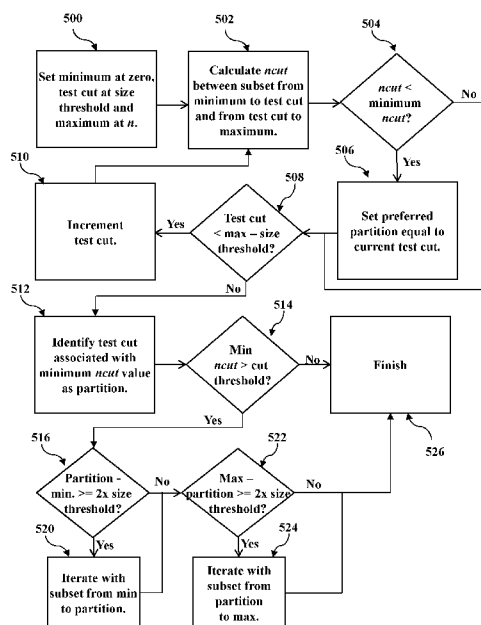


FIG. 5

(57) Abstract: Aspects of the disclosed technology can be used to implement methods in which a co-occurrence matrix of test types can be transformed through a process which includes sorting based on an eigenvector corresponding to a non-zero eigenvalue, and the transformed matrix can then be used to efficiently identify types of tests with high co-occurrence. Alternative approaches which use modified k-means clusters are also possible and could be applied in similar contexts as approaches using eigenvector sorting.

TEST PANEL ANALYSIS

FIELD

- [0001] The disclosed technology pertains to identifying clusters having high co-occurrence, such as groups of diagnostic tests that are performed together at high frequency.

BACKGROUND

- [0002] Often, when blood or another body fluid is analyzed, it may typically be subjected to multiple different tests. For some diagnostic tasks or specific diseases there may be recommendations of pre-defined groups of tests (“panels”) that should be run together to ascertain a more complete picture of a patient’s condition. However, it is possible that clinicians may have their own preferred panels that could differ from the recommendations, or that they may diverge from recommended panels by ordering tests in an ad-hoc or non-systematic manner. Additionally, there may not always be applicable panel recommendations, and so even a clinician who consistently follows such recommendations when they are available may at times make their own idiosyncratic test orders simply as a result of recommended panels being unavailable. This can cause various problems, such as waste in the event a clinician orders tests that are redundant for each other.

SUMMARY

- [0003] There is a need for improved technology for identifying groups of tests that may be run together with a high frequency. It may thus be an object of some embodiments to provide a method that could comprise steps such as obtaining a set of co-occurrence data for each of a plurality of test types, defining a co-occurrence distribution based on co-occurrence data, generating a transformation operator (e.g., a derivative operator such as a Laplacian matrix) based on the co-occurrence distribution, generating a sorting construct based on the transformation operator, generating an evaluation distribution based on sorting the co-occurrence distribution with the sorting construct, and generating a set of co-occurrence

clusters for the plurality of types of test based on the evaluation distribution. In some embodiments, this objective may be fulfilled by the subject matter of the independent claims, wherein further embodiments are incorporated in the dependent claims.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0004] The drawings and detailed description that follow are intended to be merely illustrative and are not intended to limit the scope of the invention as contemplated by the inventors.
- [0005] FIG. 1 is an architecture which may be used in some embodiments.
- [0006] FIG. 2 is an exemplary co-occurrence distribution in the form of a matrix.
- [0007] FIG. 3 is a flowchart showing a process which may be used in some embodiments to derive an evaluation distribution from a co-occurrence distribution.
- [0008] FIG. 4 is an exemplary evaluation distribution in the form of a matrix.
- [0009] FIG. 5 is an exemplary process that may be used to automatically identify co-occurrence clusters in some embodiments.
- [0010] FIG. 6 is an exemplary process that may be used to automatically identify co-occurrence clusters in some embodiments.

DETAILED DESCRIPTION

- [0011] In light of the above, it could be beneficial to be able to identify tests which are often ordered together both to address problems such as waste, as well as for other purposes such as identifying emerging trends in testing. However, with conventional approaches it has not been feasible to identify these types of patterns. According to a first aspect some embodiments may include a method comprising steps such as obtaining a set of co-occurrence data for each of a plurality of types of tests performed on patient samples, defining a co-occurrence distribution based

on the co-occurrence data, determining a transformation operator based on the co-occurrence distribution, determining a sorting construct based on the transformation operator, generating an evaluation distribution based on sorting the co-occurrence distribution using the sorting construct determined based on the transformation operator, and generating a set of co-occurrence clusters for the plurality of types of tests based on the evaluation distribution.

[0012] According to a second aspect, in some embodiments such as described in the context of the first aspect, the evaluation distribution may be a matrix in which each type of test from the plurality of types of tests corresponds to one row and one column. In some such embodiments, generating the set of co-occurrence clusters based on the evaluation distribution may comprise displaying a representation of the evaluation distribution in which each off-diagonal element of the evaluation distribution is displayed in a cell having a color determined based on relative frequency of co-occurrences for tests of the type corresponding to that off-diagonal element's column with tests of the type corresponding to that off-diagonal element's row. In such embodiments, generating the set of co-occurrence clusters based on the evaluation distribution may also comprise receiving input from a user, the input indicating one or more sections of the evaluation distribution which should be grouped together into co-occurrence clusters.

[0013] According to a third aspect, in some embodiments such as described in the context of the first aspect, generating the set of co-occurrence clusters based on the evaluation distribution may comprise performing a partitioning process on a defined portion of the evaluation distribution. In some embodiments where it is present, such a partitioning process may comprise, for each of a set of one or more types of test taken from the types of tests in the defined portion of the evaluation distribution, determining a connection value associated with partitioning between that type of test and the next type of test from the defined portion of the evaluation distribution. In such embodiments, the partitioning process may further comprise

identifying a partition associated with a lowest determined connection value as the partition to apply to the defined portion of the evaluation distribution.

[0014] According to a fourth aspect, in some embodiments which comprise a partitioning process as described in the context of the third aspect, the partitioning process may comprise, after identifying the partition to apply to the defined portion of the evaluation distribution, determining whether to further partition any sub-portion of the defined portion of the evaluation distribution defined based on the identified partition. In such embodiments, the partitioning process may also comprise, for each sub-portion of the defined portion of the evaluation distribution where a determination is made to further partition that sub-portion, performing the partitioning process with that sub-portion as the defined portion of the evaluation distribution.

[0015] According to a fifth aspect, in some embodiments which comprise performing the partitioning process with sub-portions of the defined portion of the evaluation distribution such as described in the context of the fourth aspect, the partitioning process may comprise, for each sub-portion of the defined portion of the evaluation distribution where a determination is made to further partition that sub-portion, before determining connection values associated with partitions in that sub-portion, performing a set of steps for that sub-portion. In some embodiments, such a set of steps may comprise generating a transformation operator based on that sub-portion, generating a sorting construct based on the transformation operator generated based on that sub-portion, and sorting that sub-portion with the sorting construct generated based on the transformation operator generated based on that sub-portion.

[0016] According to a sixth aspect, in some embodiments of the types described in the context of either of the third or fourth aspect, determining whether to further partition any sub-portion of the defined portion of the evaluation distribution may comprise comparing a connectedness threshold with a connectedness value

between that sub-portion and another sub-portion of the defined portion of the evaluation distribution defined based on the identified partition. In such embodiments, determining whether to further partition any sub-portion of the defined portion of the evaluation distribution may further comprise comparing a size of that sub-portion with a cluster size threshold.

[0017] According to a seventh aspect, in some embodiments of the type described in the context of the sixth aspect, a connectedness value may be determined using an equation that combines connectedness metrics for sub-portions of the defined portion of the evaluation distribution defined based on the identified partition.

[0018] According to a eighth aspect, in some embodiments of the type described in the context of the third aspect which comprise a partitioning process comprising performing acts for each of a set of one or more types of tests taken from the types of tests in the defined portion of the evaluation distribution, the set of one or more types of tests taken from the types of tests in the defined portion of the evaluation distribution may comprise each type of test in the defined portion of the evaluation distribution.

[0019] According to an ninth aspect, in some embodiments such as described in the context of the first aspect, the co-occurrence data may comprise, for each of the plurality of types of tests as a subject test type, for each other type of test from the plurality of types of tests, a number of times tests of the subject test type were included in a single order with tests of that other test type. In some such embodiments, the co-occurrence distribution may be a symmetrical co-occurrence matrix in which each type of test corresponds to one row and one column and each off-diagonal element in the co-occurrence matrix may represent the number of times tests having the test type corresponding to that off-diagonal element's column were included in a single order with tests having the test type corresponding to that off-diagonal element's row. Further in some such embodiments, the transformation operator generated based on the co-occurrence

distribution may be a Laplacian matrix, the sorting construct generated based on the transformation operator may be a first nonzero eigenvector of the Laplacian matrix, and the evaluation distribution may be a matrix in which each type of test from the plurality of types of tests corresponds to one row and one column.

[0020] Corresponding systems comprising one or more computers configured by computer executable instructions stored on non-transitory computer readable media to perform steps of methods described in any of the preceding embodiments, as well as non-transitory computer readable media storing instructions for performing steps of method described in any of the preceding embodiments, as well as systems and machines configured to perform methods as described in the context of any of the preceding embodiments could also be implemented without undue experimentation by those of ordinary skill in the art based on this disclosure. Accordingly, the preceding description of potential embodiments and aspects should be understood as being illustrative only, and should not be treated as limiting.

[0021] Turning now to the figures, FIG. 1 shows a schematic diagram of an exemplary information system 100 for identifying clusters of tests which have a high frequency of co-occurrence. The exemplary information system 100 may be configured to receive assay data from one or more lab instruments 102 104 106. This assay data may include information such as sample IDs of samples that tests are performed on, as well as the tests that were performed. As shown in figure 1, such an information system 100 may include one or more computing servers 108, and one or more memories 110 which may be used, respectively to process and store information received from the one or more lab instruments 102 104 106. Additionally, in some embodiments an information system 100 such as shown in figure 1 may also include or be in communication with a display 112 which could be used to provide either intermediate or final results of the information system's processing to a user.

[0022] Of course, it should be understood that the description above of an information system 100 obtaining assay data directly from lab instruments 102 104 106 is intended to be illustrative only, and should not be treated as limiting on the scope of protection provided by this document or any other document claiming the benefit of this disclosure. For example, in some embodiments, either in addition to, or as an alternative to, obtaining assay data directly from laboratory instruments, an information system 100 may obtain such data from either laboratory information systems 114 or hospital information systems 116. Similarly, it should be understood that while figure 1 illustrates various laboratory instruments 102 104 106 as being connected to the information system 100 via a shared network (e.g., a common LAN), it is possible that, in some embodiments, an information system may collect assay data from laboratory instruments which are not so interconnected (e.g., instruments located at different laboratories that do not share a common network). Similarly, while figure 1 illustrates only a single laboratory information system 114 and hospital information system 116, it is possible that some embodiments may collect assay data from multiple laboratory information systems, and/or multiple hospital information systems. Accordingly, while some embodiments may follow the architecture shown in figure 1, that architecture should be seen as illustrative only, and should not be treated as limiting.

[0023] Turning now to FIG. 2, that figure shows a co-occurrence distribution represented as a co-occurrence matrix for generic tests such as could be run on lab instruments 102 104 106 of figure 1. In that matrix, the diagonal entries indicate the number of times a particular test was ordered (e.g., test T1 was ordered 123,378 times, test T2 was ordered 103,661 times, etc.). The off-diagonal entries then illustrate the number of times different tests were ordered together. For example, number 1,913 in the second column of the first row and the second row of the first column indicates that tests T1 and T2 were ordered together 1,913 times. As will be apparent, a co-occurrence distribution such as the co-occurrence matrix shown in

figure 2 may be populated in a variety of manners. For example, in some embodiments an information system 100 such as shown in figure 1 may populate a matrix such as shown in figure 2 by pulling a subset of the data stored in its memory/memories 110 (e.g., co-occurrence data from the preceding two weeks for the most frequently ordered types of tests). Alternatively, in some embodiments, all assay data from an information system's memory/memories 110 could be used to populate a co-occurrence distribution, thereby providing a more comprehensive view of the data available to the organization maintaining the information system 100. Of course, in some embodiments a user of the information system 100 may be able to specify the data that should be used to populate a co-occurrence distribution, thereby providing flexibility for different population approaches to be used for different purposes (e.g., populating with a recent subset of data for identifying current practices, versus populating with a subset corresponding to a particular past time period to identify differential effects that the policies in place during the historical and more recent time periods may have had). Accordingly, the above described approaches to populating a co-occurrence distribution such as the co-occurrence matrix shown in figure 2 should be understood as being illustrative only, and should not be treated as limiting.

[0024] Turning now to FIG. 3, that figure shows a process that, in some embodiments, may be used to derive an evaluation distribution from a co-occurrence distribution such as the co-occurrence matrix shown in figure 2. Initially in the process of figure 3, a transformation operator (e.g., a derivative operator, such as a Laplacian matrix) will be determined 300 based on the co-occurrence matrix. This may be done by, to use the creation of a Laplacian matrix (also referred to herein as a "Laplacian") from a co-occurrence matrix as an example, by creating a diagonal matrix D , in which each element D_{ii} is the sum of the elements from the i^{th} row of the co-occurrence matrix, and then subtracting the co-occurrence matrix to obtain a new matrix $L = D - W$. After the transformation operator has been determined 300, the process of figure 3 continues with generating 302 a sorting construct

which, in some embodiments, may be the first non-zero eigenvector (e.g., the eigenvector corresponding to the lowest non-zero eigenvalue) of the transformation operator (e.g., the first non-zero eigenvector of the Laplacian, in embodiments where the operator is a Laplacian matrix).

[0025] With the sorting construct available, in some embodiments, a process such as shown in figure 3 could continue with sorting 304 the co-occurrence distribution using the sorting construct. In some embodiments where the co-occurrence distribution is a co-occurrence matrix and the sorting construct is the first non-zero eigenvalue of the co-occurrence matrix's Laplacian, this sorting 304 could be done by recording the position of each element in the eigenvector, sorting the eigenvector while tracking each element's position, using the relationships between the original and final positions of the eigenvector's elements to create a mapping from the original to the final ordering (e.g., a function f that, for an input integer n representing a position in the sorted eigenvector, returns an integer output showing the position that the n^{th} element in the sorted eigenvector occupied in the unsorted eigenvector), and applying that mapping to the test correlation matrix to obtain a sorted test correlation matrix (e.g., creating a new matrix in which each element E_{ij} is equal to element $f(i)f(j)$ in the unsorted co-occurrence matrix). An example of a sorted matrix such as could be obtained by applying the above steps to the matrix of figure 2 is illustrated in figure 4. Please note that while the diagonal entries in figure 4 are 0 (representing that, in an adjacency graph representation of test co-occurrence, vertices would have edges connecting each other but not themselves), this preferably will not impact the analysis since the co-occurrences are reflected by the off-diagonal elements.

[0026] After an evaluation distribution (e.g., a matrix such as shown in figure 4) has been derived, in some embodiments the evaluation distribution may be used to identify clusters of tests with high (or relatively high) frequency of co-occurrence. As will be apparent to those of ordinary skill in the art, there are various ways in which this clustering could be performed, and different embodiments may utilize

different approaches or combinations of approaches to clustering. For example, in some embodiments, clusters of tests with high co-occurrence may be identified by presenting a matrix representation of the evaluation distribution to a human operator and taking advantage of the fact that the human eye is generally very skilled in identifying visual patterns and groupings. In embodiments which include this type of visual clustering, there may be preparatory steps performed to facilitate identification of groups. For instance, cells in the matrix representation of the evaluation distribution may be colored to show their value (e.g., linearly or logarithmically transitioning from pure green for cells with a value of 0 to pure red to the cells with the highest values) and therefore the co-occurrence of the tests corresponding to the rows and columns of the relevant cells. Other types of preparation may also be performed in some cases. For example, in some embodiments, prior to coloring diagonal elements in a matrix representation of an evaluation distribution may be set equal to the average of their adjacent cells so that they would tend to blend in and enhance clusters rather than bisect and detract from the user's ability to identify them.

[0027] It should be understood that the above description of visually identifying clusters in a matrix representation of an evaluation distribution is intended to be illustrative only, and that other approaches may be used in some embodiments. For example, in some embodiments, a process such as shown in figure 5 may be applied to automatically identify clusters in an evaluation distribution such as the sorted analysis matrix shown in figure 4. At a high level, the process of figure 5 will segment an evaluation distribution into sub-portions (e.g., in the case of a matrix representation of the evaluation distribution, submatrices lying along the original analysis matrix's diagonal) and will then iteratively segment those sub-portions so long as various cluster size and connectivity requirements are met. In more detail, the process of figure 5 begins with setting values 500 that will be used in later processing. Specifically, the process of figure 5 will set a minimum value (used for defining the upper left corner of the upper left submatrix) at zero, a maximum

value (used in embodiments which represent an evaluation distribution in matrix form for defining the lower right corner of the lower right submatrix) at n (e.g., the number of rows/columns in an analysis matrix), and a test cut value (used to determine where to partition the evaluation distribution, such as a location along an analysis matrix's diagonal) equal to a sub-portion size threshold (a parameter defining the minimum size for a cluster, which will preferably be set by a user, but may also be set automatically as a percentage of the total number of elements in the evaluation distribution, the total number of rows/columns in a matrix representation of an evaluation distribution, or as a constant default value).

[0028] With the analysis values having been set 500, a process such as shown in figure 5 will continue with calculating 502 an *ncut* value between subsets of elements defined by the test cut value. The *ncut* value can be seen as a measure of how connected the portions of the evaluation distribution are to each other. In embodiments where the evaluation distribution takes the form of an analysis matrix such as shown in figure 4, this can be calculated using equation 1, below:

$$ncut(A, B) = \frac{c(A, B)}{c(A, A + B)} + \frac{c(A, B)}{c(B, A + B)}$$

Equation 1: *ncut* value calculation

In equation 1, A and B represent subsets of row/columns (which can be seen as interchangeable since the analysis matrix is symmetric) that would be divided by the test cut value and $c(A, B)$ is the sum of the connections between subset A and subset B . Thus, with the analysis matrix of figure 4, if the test cut value is three, then subset A would be the three rows corresponding to tests T22, T19 and T21, and subset B would be the 20 rows starting with the row representing test T10 and continuing through the row representing test T8). $c(A, B)$ could then be found by summing the relevant elements in those rows using equation 2, below:

$$c(A, B) = \sum_{i \in A, j \in B} w_{ij}$$

Equation 2: exemplary connectivity measure equation

[0029] In the process of figure 5, after the *ncut* value has been calculated 502, it is compared 504 with the smallest previously calculated *ncut* value (or, in some embodiments, if no previous *ncut* value has been calculated for the current minimum and maximum parameters, this step may be skipped). If the *ncut* value for the current test cut is less than the smallest previously calculated *ncut* value, then the current test cut can be identified 506 as the preferred partition, reflecting the fact that it is the best way (so far) identified for separating the elements defined by the maximum and minimum values. After the *ncut* value has been checked 504 and the preferred partition value updated 506 (if needed), a further check 508 can be made of whether the current test cut value is greater than the previously set maximum size minus the previously defined sub-portion size threshold. In a process such as shown in figure 5, this type of test 508 could be used to prevent needlessly checking partitions that would create sub-portions that are smaller than the previously defined threshold size. Then, if the test cut value was greater than the maximum value minus the size threshold, the test cut value could be incremented 510 and the process could return to calculating 502 the *ncut* value for the new test cut value, and this type of iteration could be repeated until the entire diagonal of the analysis matrix (less the left and right portions which were less than the size threshold) had been traversed.

[0030] After all potential partitions between the minimum and maximum values (e.g., after the diagonal of an analysis matrix from the minimum to maximum value (less thresholds) had been traversed) the test cut associated with the minimum *ncut* value would be identified 512 as the value to use for partitioning the elements

between the previously set minimum and maximum values. At this point, in some embodiments a set of one or more checks might be performed to determine if further divisions should be made in either of the sub-portions. For example, in some embodiments, a predefined connectedness threshold may have been set such that a check 514 showing the *ncut* value did not exceed that threshold would be treated as indicating that no further partitioning between the maximum and minimum values was necessary. Similarly, in some embodiments, a check 516 could be performed to confirm if the subset from the minimum value to the partition was at least twice the minimum cluster size. Then, if it was, the process could iterate 518 by partitioning that subset (e.g., by leaving the minimum unchanged, setting the maximum to the current partition value, and returning to the previously described calculation 502 of *ncut*). Similarly, in some embodiments a check 522 may be performed if the subset from the partition to the maximum value was at least twice the minimum cluster size. Then, if it was, the process could iterate 524 by partitioning that subset (e.g., by setting the minimum equal to the current partition value, leaving the maximum unchanged, and returning to the previously described calculation 502 of *ncut*). Finally, in the process of figure 5, once the various checks (e.g., the checks 514, 516 and/or 522 shown in figure 5) indicated that no further partitioning was needed, the process could finish 526, and the various subsets defined by the identified partitions could be treated as the clusters for the evaluation distribution.

[0031] Of course, it should be understood that the above description is intended to be illustrative only, and that numerous other embodiments are possible and could be implemented without undue experimentation based on this disclosure by those of ordinary skill in the art. For example, in some embodiments, rather than treating all subsets identified using partitions as separate clusters of tests, a further check could be performed on each subset testing whether the connectedness of its elements relative to elements in the rest of the evaluation distribution (e.g., using the *ncut* calculation described above, potentially, but not necessarily, with a

different threshold than was used for initially determining whether to continue iteration) was sufficient to justify treating it as a cluster that might be worthy of further study. Similarly, in some embodiments, rather than testing for size thresholds before deciding whether to iterate, or limiting iteration to sections of an evaluation distribution limited by a cluster size threshold, testing for size thresholds may be performed subsequently – such as in determining whether a test cut should be treated as a partition used in further analysis.

[0032] Other types of variations, including variations with diverge from the basic framework depicted in figure 3, are also possible. As an example, consider the process illustrated in figure 6, which may be used in some embodiments. Initially, in the process of figure 6, a determination 600 would be made of a target number of clusters that the data should be organized into (e.g., by asking a user to input a target number of clusters), and a corresponding determination 602 would be made of that number of characteristic constructs (e.g., eigenvectors of an operator determined based on the co-occurrence matrix, such as a co-occurrence matrix's Laplacian). For instance, in some embodiments, if the first determination 600 was that the data should be organized into k clusters, then the second determination 602 would preferably be of the first k eigenvectors for the Laplacian. In some embodiments, a determination 604 may also be made of a randomization value (e.g., by setting the randomization to a default value, such as 1, or to a value proportionate to the maximum elements in the distribution, such as 10% of the absolute value of an analysis matrix's largest element) that could be used to reduce the risk that the clustering algorithm would fall into a sub-optimal local minimum.

[0033] After the target number of characteristic constructs had been determined 602 (and, in some embodiments, potentially before determination 604 of an initial randomization value), a process such as shown in figure 6 may proceed with generating 606 an analysis space based on those characteristic constructs. In some embodiments where the characteristic constructs are eigenvectors of a co-occurrence matrix's Laplacian, this may be done by assembling the eigenvectors

into an analysis matrix in which each eigenvector was a column of the analysis matrix. An initial cluster assignment could then be set 608 for the analysis space. For example, in some embodiments, this could be done by treating each row of an analysis matrix as a point in k-dimensional space (where k is the previously determined target cluster number), and then randomly assigning each of those points to one of k clusters. Alternatively, in some embodiments, initial clusters could be set by randomly choosing k rows of an analysis matrix and treating them as cluster centroids. Other approaches to initially assigning clusters, such as requesting a user to make a best guess of how rows in an analysis matrix should be grouped into clusters are also possible, and could be implemented without undue experimentation by those of ordinary skill in the art in light of this disclosure.

[0034] In some embodiments which implement a method such as depicted in figure 6, after the initial values had been set/determined, points in the analysis space could be assigned 610 to clusters based on the then current centroids and the randomization values. This could be done, for example, by performing a calculation that treats each row in an analysis matrix as a point in k dimensional space, and then measures the distance (e.g., the Euclidian distance, though other distance measures may be used in some embodiments) between that point and the locations of the centroids for the then current clusters. The points (e.g., the rows in the analysis matrix) could then be assigned to the clusters with the closest centroids. Additionally, in some embodiments the distances may be modified using the randomization value. For example, for a point p with distances $d_1, d_2, \dots, d_k$ from the k centroids, each of the distances might be randomly modified based on the randomization value (e.g., multiplied by the product of the randomization value and a random number between -1 and 1) before point p was assigned to a cluster. In this way, some embodiments may reduce the risk that their cluster assignments will become trapped at a local minimum, since the additional

randomization might introduce enough noise to break out of a local minimum once it was entered.

[0035] After the points in the analysis space had been assigned 610 to clusters, in some embodiments, a method such as depicted in figure 6 could check to see if that assignment was different from the preceding assignment. For instance, the initial cluster assignment was set 608 randomly, the check 612 would examine whether any of the points were in different clusters than the ones to which they had initially been randomly assigned. If any assignments had changed, then an update 614 could be performed by reducing the randomization value (e.g., dividing it by 10) and recalculating the cluster centroids using the new cluster assignments. The process could then iterate by repeating the assignment 610 step, checking for changes 612, and continuing until the assignments stabilized and no changes were detected. At this point, the clustering could be deemed to be complete and the underlying data could be assigned 616 to clusters based on the clustering of the points in the analysis space (e.g., if row 1 of an analysis matrix representation of k dimensional analysis space was assigned to cluster 1, then the test type corresponding to row 1 of the underlying co-occurrence matrix could be assigned to cluster 1, if row 2 of the analysis matrix was assigned to cluster 3 then the test type corresponding to row 2 of the underlying co-occurrence matrix could be assigned to cluster 3, etc.).

[0036] Further variations on, and features for, the inventors' technology will be immediately apparent to, and could be practiced without undue experimentation by, those of ordinary skill in the art in light of this disclosure. For example, in some embodiments which utilize k-means clustering as described in the context of figure 6, the k-means clustering may be applied directly to a co-occurrence matrix (or other representation of a co-occurrence distribution), rather than to a derived construct such as rows in an analysis matrix. Similarly, in some embodiments identification of clusters using k-means clustering may be part of a larger process in which a user would specify a target cluster number, be presented with that

number of clusters determined using k-means clustering, and then be able to repeat with a new target cluster number and compare the results to determine final clustering for the test types. Accordingly, instead of limiting the protection accorded by this document, or by any document which is related to this document, to the material explicitly disclosed herein, the protection should be understood to be defined by the claims, if any, set forth herein or in the relevant related document when the terms in those claims which are listed below under the label “Explicit Definitions” are given the explicit definitions set forth therein, and the remaining terms are given their broadest reasonable interpretation as shown by a general purpose dictionary. To the extent that the interpretation which would be given to such claims based on the above disclosure is in any way narrower than the interpretation which would be given based on the “Explicit Definitions” and the broadest reasonable interpretation as provided by a general purpose dictionary, the interpretation provided by the “Explicit Definitions” and broadest reasonable interpretation as provided by a general purpose dictionary shall control, and the inconsistent usage of terms in the specification or priority documents shall have no effect.

[0037] Explicit Definitions

[0038] When appearing in the claims, a statement that something is “based on” something else should be understood to mean that something is determined at least in part by the thing that it is indicated as being “based on.” When something is required to be completely determined by a thing, it will be described as being “based exclusively on” the thing.

[0039] When used in the claims, “determining” should be understood to refer generating, selecting, defining, calculating or otherwise specifying something. For example, to obtain an output as the result of analysis would be an example of “determining” that output. As a second example, to choose a response from a list of possible responses would be a method of “determining” a response. As a third example, to

identify data received from an external source (e.g., a microphone) as being a thing would be an example of “determining” the thing.

[0040] When used in the claims a “means for automatically identifying co-occurrence clusters from tests performed on one or more laboratory instruments” should be understood as a means plus function limitation as provided for in 35 U.S.C. § 112(f), in which the function is “automatically identifying co-occurrence clusters from tests performed on one or more laboratory instruments” and the corresponding structure is a computer configured to perform processes as illustrated in figure 5 and described in the corresponding text.

CLAIMS

1. A method comprising:
 - a) obtaining a set of co-occurrence data for each of a plurality of types of tests performed on patient samples;
 - b) defining a co-occurrence distribution based on the set of co-occurrence data;
 - c) determining a transformation operator based on the co-occurrence distribution;
 - d) determining a sorting construct based on the transformation operator;
 - e) generating an evaluation distribution based on sorting the co-occurrence distribution using the sorting construct determined based on the transformation operator; and
 - f) generating a set of co-occurrence clusters for the plurality of types of tests based on the evaluation distribution.
2. The method of claim 1, wherein:
 - a) the evaluation distribution is a matrix in which each type of test from the plurality of types of tests corresponds to one row and one column;
 - b) generating the set of co-occurrence clusters based on the evaluation distribution comprises
 - i) displaying a representation of the evaluation distribution in which each off-diagonal element is displayed in a cell having a color determined based on relative frequency of co-occurrences for tests of the type corresponding to that off-diagonal element's column with tests of the type corresponding to that off-diagonal element's row; and
 - ii) receiving input from a user, the input indicating one or more sections of the evaluation distribution which should be grouped together into co-occurrence clusters.
3. The method of claim 1, wherein generating the set of co-occurrence clusters based on the evaluation distribution comprises performing a partitioning process on a defined portion of the evaluation distribution, wherein the partitioning process comprises:
 - a) for each of a set of one or more types of tests taken from the types of tests in the defined portion of the evaluation distribution, determining a connection value associated with

- partitioning between that type of test and the next type of test from the defined portion of the evaluation distribution; and
- b) identifying a partition associated with a lowest determined connection value as the partition to apply to the defined portion of the evaluation distribution.
4. The method of claim 3, wherein the partitioning process comprises:
- a) after identifying the partition to apply to the defined portion of the evaluation distribution, determining whether to further partition any sub-portion of the defined portion of the evaluation distribution defined based on the identified partition; and
- b) for each sub-portion of the defined portion of the evaluation distribution where a determination is made to further partition that sub-portion, performing the partitioning process with that sub-portion as the defined portion of the evaluation distribution.
5. The method of claim 4, wherein the partitioning process comprises, for each sub-portion of the defined portion of the evaluation distribution where a determination is made to further partition that sub-portion, before determining connection values associated with partitions in that sub-portion:
- a) determining a transformation operator based on that sub-portion;
- b) determining a sorting construct based on the transformation operator determined based on that sub-portion; and
- c) sorting that sub-portion with the sorting construct determined based on the transformation operator determined based on that sub-portion.
6. The method of claim 4, wherein determining whether to further partition any sub-portion of the defined portion of the evaluation distribution comprises:
- a) comparing a connectedness value between that sub-portion and another sub-portion of the defined portion of the evaluation distribution defined based on the identified partition with a connectedness threshold; and
- b) comparing a size of that sub-portion with a cluster size threshold.

7. The method of claim 6 wherein the connectedness value is determined using an equation that combines connectedness metrics for sub-portions of the defined portion of the evaluation distribution defined based on the identified partition.
8. The method of claim 3 wherein the set of one or more types of tests taken from the types of tests in the defined portion of the evaluation distribution comprises each type of test in the defined portion of the evaluation distribution.
9. The method of claim 1, wherein:
 - a) the co-occurrence data comprises, for each of the plurality of types of tests as a subject test type:
 - i) for each other type of test from the plurality of types of tests, a number of times tests of the subject test type were included in a single order with tests of that other test type;
 - b) the co-occurrence distribution is a symmetrical co-occurrence matrix, wherein:
 - i) each type of test corresponds to one row and one column in the co-occurrence matrix;
 - ii) each off-diagonal element in the co-occurrence matrix represents the number of times tests having the test type corresponding to that off-diagonal element's column were included in a single order with tests having the test type corresponding to that off-diagonal element's row;
 - c) the transformation operator generated based on the co-occurrence distribution is a Laplacian matrix and the sorting construct generated based on the transformation operator is a first nonzero eigenvector of the Laplacian matrix; and
 - d) the evaluation distribution is a matrix in which each type of test from the plurality of types of tests corresponds to one row and one column.
10. A machine or system configured to perform the method as claimed in any of the preceding claims 1-9.

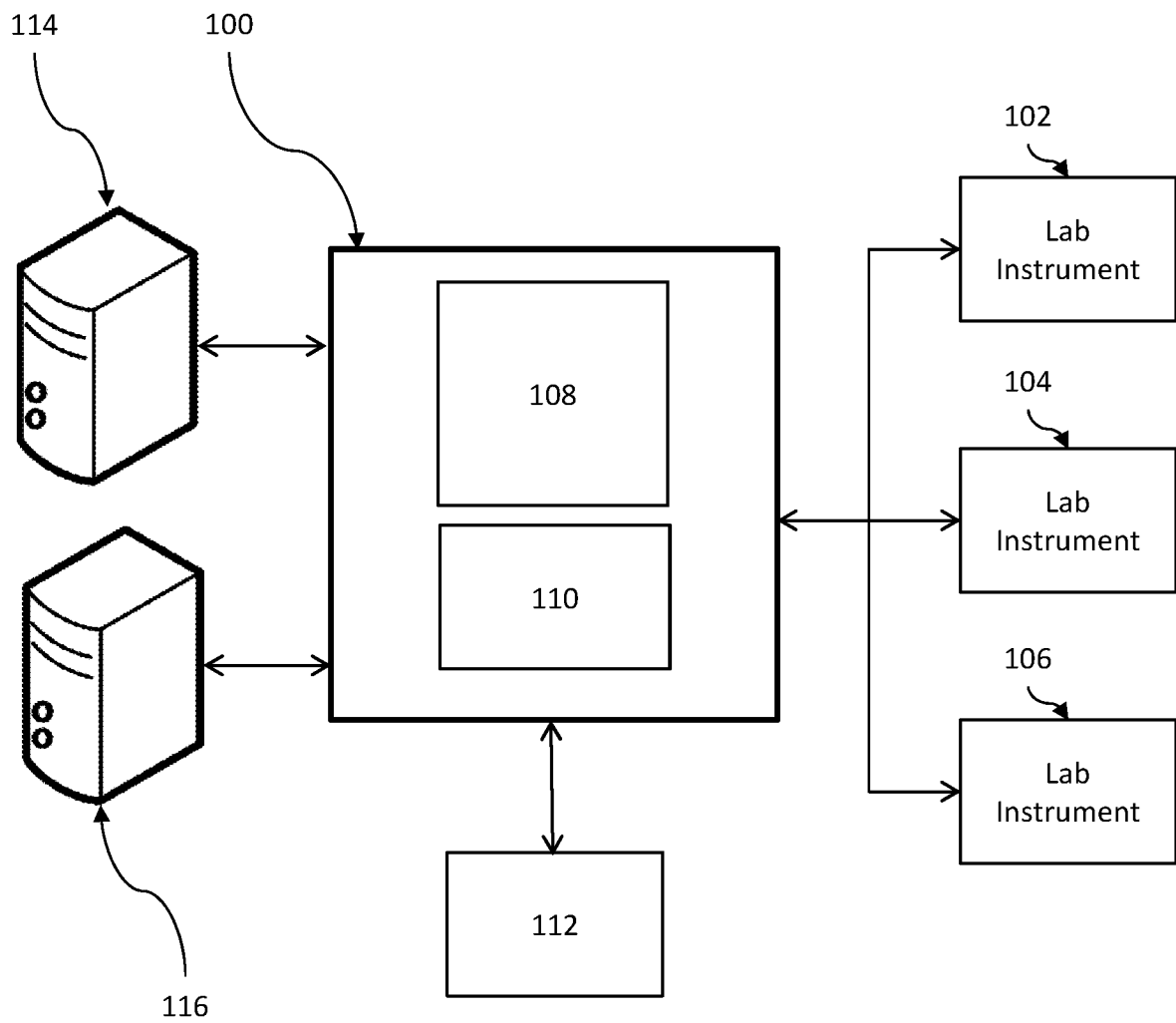
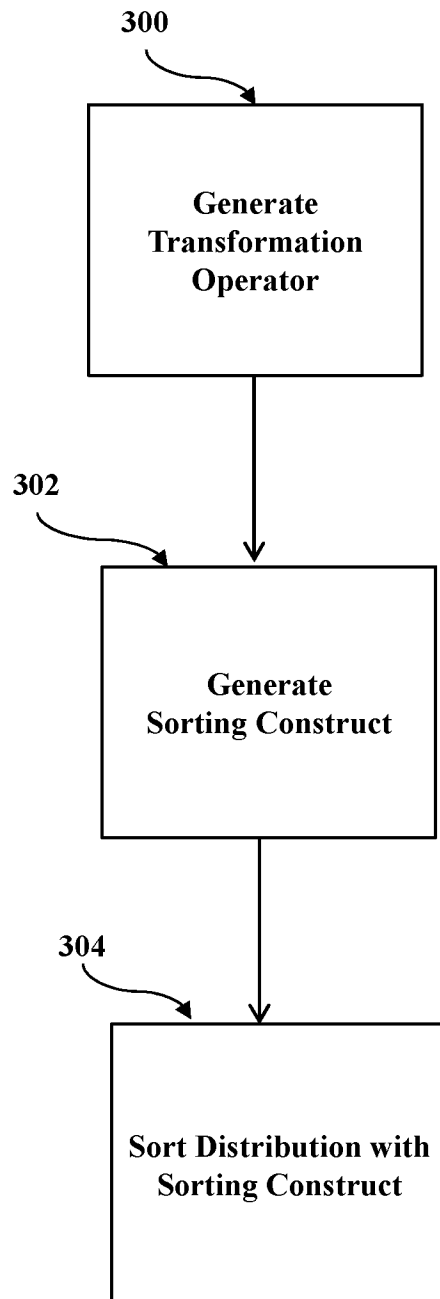


FIG. 1

	TSH3	TnIDx	TSH	FRT4	Ferritin	VitB12	PSA-Hyb	BNP2	FOLW	AFP	VitD	FTSH2	CK-MB	FT3	PTH	TPOAb	CEA2	HCG5	InhibinA	TotT4	uE3	HCG5d	hFSH
TSH3	121340	1852	143	44745	20684	14977	6494	125	11489	1002	6234	133	298	14057	1162	3774	2410	1061	17	2340	21	89	3956
TnIDx	1852	106866	1755	1407	666	723	287	936	492	116	221	996	18193	453	149	85	140	362	5	164	7	34	105
TSH	143	1755	97425	30654	5228	11920	5284	231	4294	1417	6177	614	706	4615	935	1019	1148	690	25	6547	27	63	1724
FRT4	44745	1407	30654	91876	11386	13222	4296	197	8735	1339	6044	10718	524	21784	1236	4652	2130	931	27	5000	34	71	3363
Ferritin	20684	666	5228	11386	74645	21796	3558	122	15380	1008	6120	2202	266	3434	1869	1156	1631	908	6	1556	9	67	1720
VitB12	14977	723	11920	13222	21796	65064	3626	162	32195	644	7569	4515	308	3263	1235	1265	1235	665	4	3182	11	22	1572
PSA-Hyb	6494	287	5284	4296	3558	3626	56472	96	2168	1902	1698	1982	123	1323	435	420	2472	519	3	1449	8	8	698
BNP2	125	936	231	197	122	162	96	43871	90	29	47	99	267	57	36	8	35	56	11	44	12	19	40
FOLW	11489	492	4294	8735	15380	32195	2168	90	36962	564	4005	1721	208	2036	869	925	979	604	3	2004	6	8	1153
AFP	1002	116	1417	1339	1008	644	1902	29	564	35955	201	62	49	339	112	166	4472	715	13313	566	13688	9000	311
VitD	6234	221	6177	6044	6120	7569	1698	47	4005	201	34050	1729	89	1666	2046	841	406	251	0	924	6	3	946
FTSH2	133	996	614	10718	2202	4515	1982	99	1721	62	1729	32171	549	2857	159	643	134	159	6	3682	11	30	585
CK-MB	298	18193	706	524	266	308	123	267	208	49	89	549	28699	144	52	16	80	104	1	111	0	8	35
FT3	14057	453	4615	21784	3434	3263	1323	57	2036	339	1666	2857	144	24895	685	2540	676	531	9	2533	15	21	1604
PTH	1162	149	935	1236	1869	1235	435	36	869	112	2046	159	52	685	21370	596	164	176	3	381	6	1	457
TPOAb	3774	85	1019	4652	1156	1265	420	8	925	166	841	643	16	2540	596	18604	289	180	4	640	20	9	763
CEA2	2410	140	1148	2130	1631	1235	2472	35	979	4472	406	134	80	676	164	289	18527	546	5	512	5	6	463
HCG5	1061	362	690	931	908	665	519	56	604	715	251	159	104	531	176	180	546	15370	46	334	43	49	829
InhibinA	17	5	25	27	6	4	3	11	3	13313	0	6	1	9	3	4	5	46	14411	0	13341	9419	9
TotT4	2340	164	6547	5000	1556	3182	1449	44	2004	566	924	3682	111	2533	381	640	512	334	0	13961	3	9	914
uE3	21	7	27	34	9	11	8	12	6	13688	6	11	0	15	6	20	5	43	13341	3	13931	8992	13
HCG5d	89	34	63	71	67	22	8	19	8	9000	3	30	8	21	1	9	6	49	9419	9	8992	13011	17
hFSH	3956	105	1724	3363	1720	1572	698	40	1153	311	946	585	35	1604	457	763	463	829	9	914	13	17	12236

FIG. 2

**FIG. 3**

	BNP2	CK.MB	TnIDx	fTSH2	TSH	FRT4	VitB12	TSH3	FOLW	FT3	Ferritin	VitD	TotT4	PTH	hFSH	TPOAb	PSA.Hyb	HCG5	CEA2	AFP	uE3	InhibinA	HCG5d
BNP2	0	267	936	99	231	197	162	125	90	57	122	47	44	36	40	8	96	56	35	29	12	11	19
CK.MB	267	0	18193	549	706	524	308	298	208	144	266	89	111	52	35	16	123	104	80	49	0	1	8
TnIDx	936	18193	0	996	1755	1407	723	1852	492	453	666	221	184	149	105	85	287	362	140	116	7	5	34
fTSH2	99	549	996	0	614	10718	4515	133	1721	2857	2202	1729	3682	159	585	643	1982	159	134	62	11	6	30
TSH	231	706	1755	614	0	30654	11920	143	4294	4615	5228	6177	6547	935	1724	1019	5284	690	1148	1417	27	25	63
FRT4	197	524	1407	10718	30654	0	13222	44745	8735	21784	11386	6044	5000	1236	3363	4652	4296	931	2130	1339	34	27	71
VitB12	162	308	723	4515	11920	13222	0	14977	32195	3263	21796	7569	3182	1235	1572	1265	3626	665	1235	644	11	4	22
TSH3	125	298	1852	133	143	44745	14977	0	11489	14057	20684	6234	2340	1162	3956	3774	6494	1061	2410	1002	21	17	89
FOLW	90	208	492	1721	4294	8735	32195	11489	0	2036	15380	4005	2004	869	1153	925	2168	604	979	564	6	3	8
FT3	57	144	453	2857	4615	21784	3263	14057	2036	0	3434	1666	2533	685	1604	2540	1323	531	676	339	15	9	21
Ferritin	122	266	666	2202	5228	11386	21796	20684	15380	3434	0	6120	1556	1869	1720	1156	3558	908	1631	1008	9	6	67
VitD	47	89	221	1729	6177	6044	7569	6234	4005	1666	6120	0	924	2046	946	841	1698	251	406	201	6	0	3
TotT4	44	111	164	3682	6547	5000	3182	2340	2004	2533	1556	924	0	381	914	640	1449	334	512	566	3	0	9
PTH	36	52	149	159	935	1236	1235	1162	869	685	1869	2046	381	0	457	596	435	176	164	112	6	3	1
hFSH	40	35	105	585	1724	3363	1572	3956	1153	1604	1720	946	914	457	0	763	698	829	463	311	13	9	17
TPOAb	8	16	85	643	1019	4652	1265	3774	925	2540	1156	841	640	596	763	0	420	180	289	166	20	4	9
PSA.Hyb	96	123	287	1982	5284	4296	3626	6494	2168	1323	3558	1698	1449	435	698	420	0	519	2472	1902	8	3	8
HCG5	56	104	362	159	690	931	665	1061	604	531	908	251	334	176	829	180	519	0	546	715	43	46	49
CEA2	35	80	140	134	1148	2130	1235	2410	979	676	1631	406	512	164	463	289	2472	546	0	4472	5	5	
AFP	29	49	116	62	1417	1339	644	1002	564	339	1008	201	566	112	311	166	1902	715	4472	0	13688	13313	9000
uE3	12	0	7	11	27	34	11	21	6	15	9	6	3	6	13	20	8	43	5	13688	0	13341	8992
InhibinA	11	1	5	6	25	27	4	17	3	9	6	0	0	3	9	4	3	46	5	13313	13341	0	9419
HCG5d	19	8	34	30	63	71	22	89	8	21	67	3	9	1	17	9	8	49	6	9000	8992	9419	0

FIG. 4

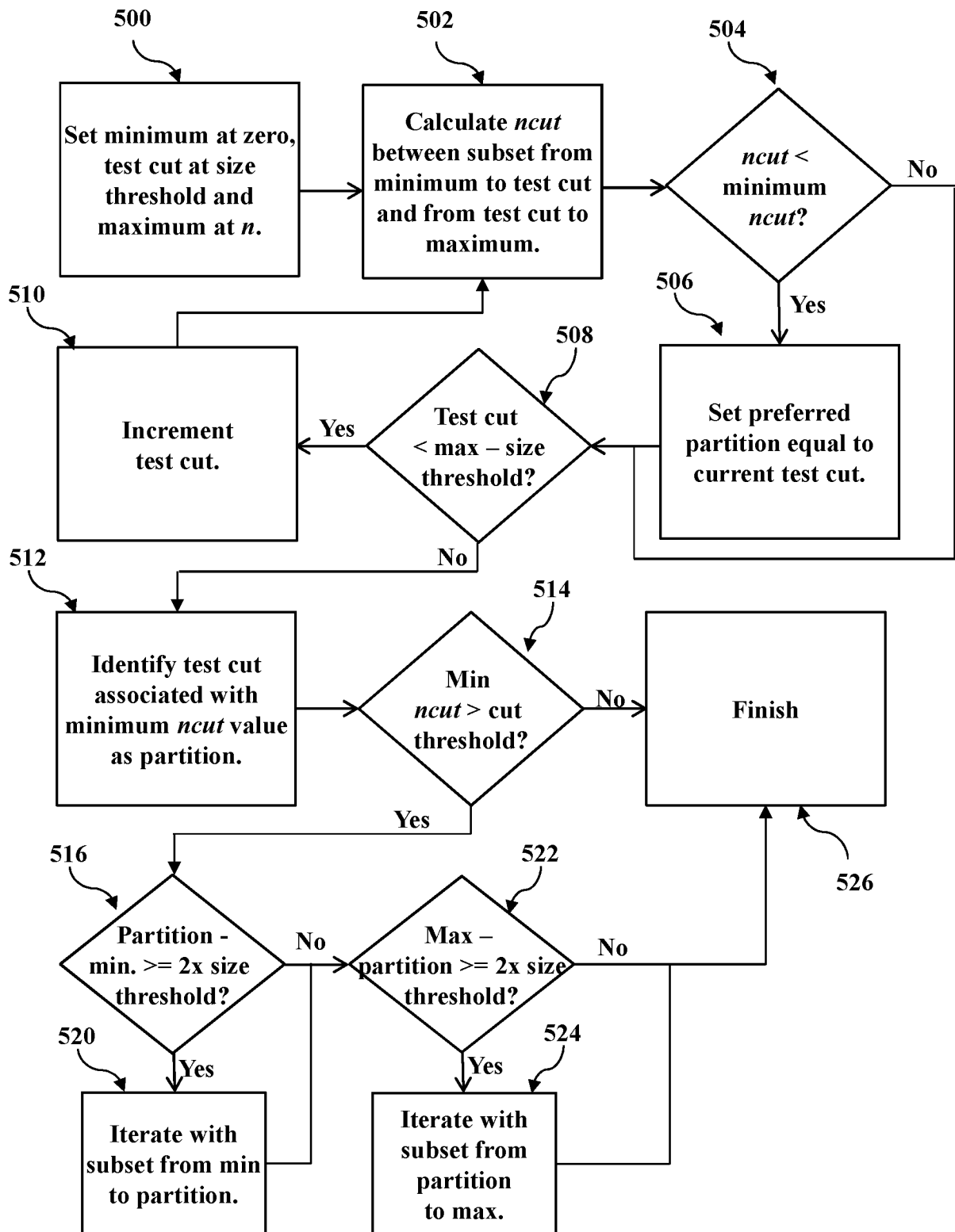


FIG. 5

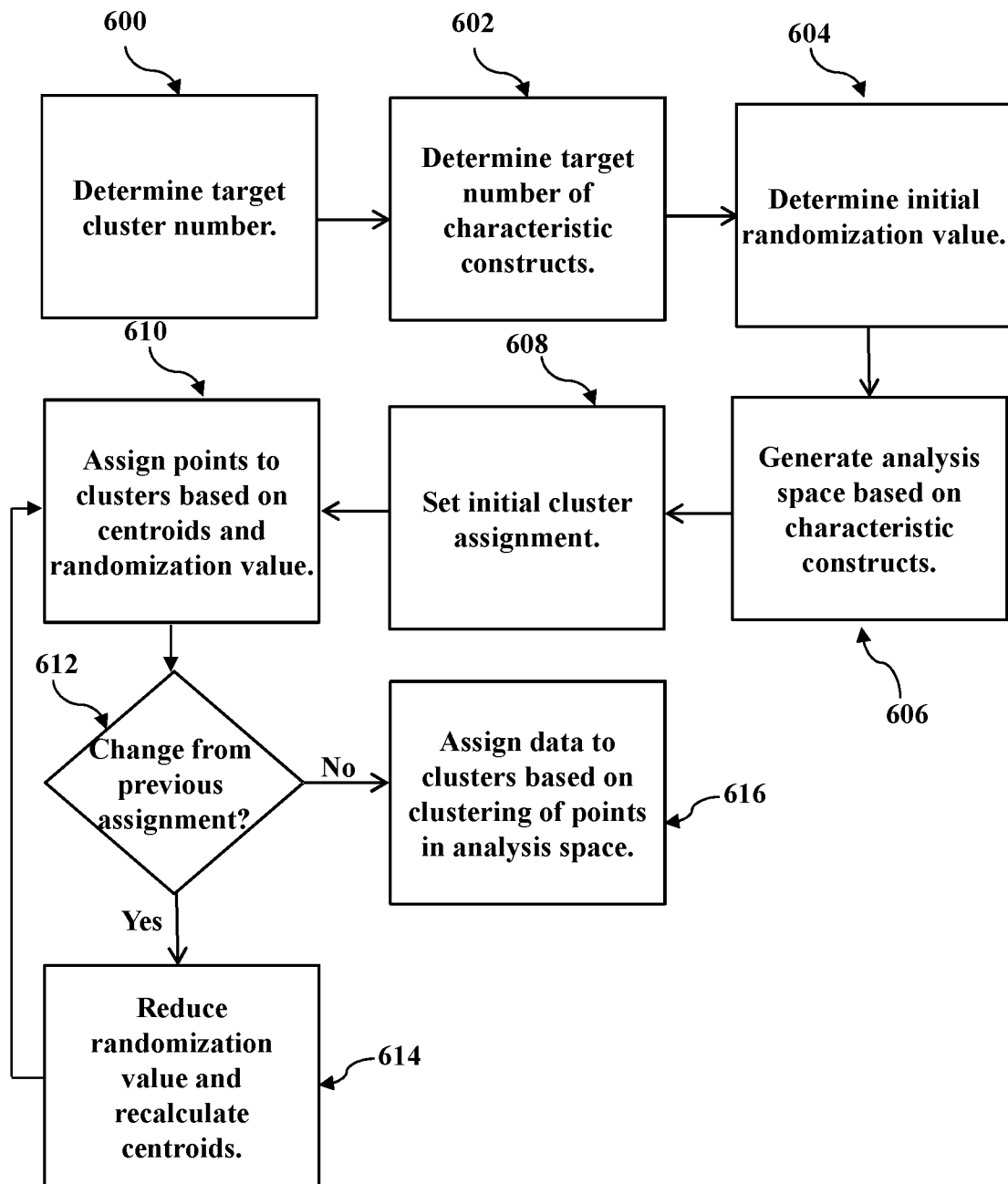


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2019/031475

A. CLASSIFICATION OF SUBJECT MATTER
INV. G16H50/70 G16H10/40
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G16H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2008/114564 A1 (IHARA MASAYOSHI [JP]) 15 May 2008 (2008-05-15) paragraph [0012] - paragraph [0022] paragraph [0237] - paragraph [0298] -----	1-10
X	US 2010/179838 A1 (BASANT NITIN [IN] ET AL) 15 July 2010 (2010-07-15) paragraph [0005] - paragraph [0038] -----	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

6 August 2019

Date of mailing of the international search report

16/08/2019

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Menschner, Philipp

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2019/031475

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2008114564 A1	15-05-2008	JP 4550882 B2	22-09-2010
		JP W02006087854 A1	07-08-2008
		US 2008114564 A1	15-05-2008
		WO 2006087854 A1	24-08-2006

US 2010179838 A1	15-07-2010	NONE	
