NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

School of Science

ITMB, *Bioinformatics*

Algorithms in Structural Bioinformatics

Intervor Project

Supervisors: Emiris Ioannis, *Chrysina Evangelia*

Postgraduates: Begetis Nikolaos, Konstantopoulos Dimitrios

Abstract: `What is Intervor?` Intervor is a software that computes a parameter free representation of macro-molecular interfaces, based on the α-complex of the atoms. `What does Intervor do?` Given two interacting partners, possibly with water molecules squeezed in-between them, Intervor computes an interface model which has the following characteristics: (i) it identifies the atoms of the partners which are in direct contact and those whose interaction is water mediated, (ii) it defines a geometric complex separating the partners, the Voronoi interface, whose geometric and topological descriptions are straightforward (surface area, number of patches, curvature), (iii) it allows the definition of the depth of atoms at the interface, thus going beyond the traditional dissection of an interface into a core and a rim. `Why do we need Intervor?` These features can be used to investigate correlations between structural parameters and key properties such as the conservation of residues, their polarity, the water dynamics at the interface, mutagenesis data, etc.

What is more, with Intervor addresses new questions such as the role played by interface atoms and in particular structural water, the flexibility issues involved in complex formation, or the specificity of recognition mechanisms

Prologue: In this project assignment readers are going to be provided with answers, verified with experiments conducted from the authors, in questions such as: (i) can one bridge the gap from atoms loosing solvent accessibility to interface pairs? (ii) Is the interface flat or curvy? (iii) Is it connected or not (does it have a multi-patch structure)? (iv) Is a connected component of the interface simply connected or not (does it have a hole)? (v) What is precisely the role played by interface structural water?

Introduction: Protein - protein recognition plays a key role in the formation of complexes which account for biological functions [(J. Janin)]. The investigation of interfaces of macro-molecular complexes is therefore central to improve our understanding of the stability and specificity of macro-molecular interactions. In conducting such investigations, one wishes to (i) improve the description of experimentally resolved complexes, and (ii) improve scoring functions used to discriminate native structures amongst putative ones.

Previous & Related Work: In this section, we review in brief the geometric pre-requisites of the Intervor model. Notice these are of major interest for the definition and the calculation of molecular surfaces and also for the characterization of pockets in molecules. Nonetheless, to understand in detail the pre-requisites one has to run in the related bibliography.

- **Van der Waals models:** *Van der Waals models* occur as the imaginary surface of the union of spherical atom surfaces defined by the so-called van der Waals radius of each atom in the molecule representation. The *van der Waals* surface enclosed volume reference is molecular volume. Both van der Waals surface and molecular volume are abstract representation of molecules, rather than "real" surfaces and volumes of molecules. *In the case of Intervor, Intervor uses the Van der Waals models in order to find atom bonds in the surface of the complex.*

- **Convex Hull:** The *convex hull* or *convex envelope* of a set X of points in the Euclidean plane is the smallest convex set that contains X. For instance, when X is a bounded subset of the plane, the convex hull may be visualized as the shape formed by a rubber band stretched around X. Formally, the convex hull may be defined as the intersection of all convex sets containing X or as the set of all convex combinations of points in X. *In the case of Intervor, Intervor uses the Convex-Hull in order to find buried of exposed atoms.*
- **Delaunay Triangulation:** A *Delaunay triangulation* for a set P of points in a plane is a triangulation DT(P) such that no point in P is inside the circumcircle of any triangle in DT(P). The Delaunay triangulation of a discrete point set P in general position corresponds to the dual graph of the Voronoi tessellation for P. Special cases include the existence of three points on a line and four points on circle. *In the case of Intervor, Intervor uses the Delaunay triangulation in order to triangulate the surface of the complex in faces.*
- **Voronoi Diagram:** a *Voronoi diagram* is a way of dividing space into a number of regions. A set of points (called seeds, sites, or generators) is specified beforehand and for each seed there will be a corresponding region consisting of all points closer to that seed than to any other. The regions are called Voronoi cells. *In relevance to the Delaunay triangulation, Intervor separates the surface of the complex in Voronoi facets, taking as center for each Voronoi facet the atoms involving in the Interface interaction.*
- **α-shapes / α-complex:** an alpha shape, or α-shape, is a family of piecewise linear simple curves in the Euclidean plane associated with the shape of a finite set of points. The alpha-shape associated with a set of points is a generalization of the concept of the convex hull, i.e. every convex hull is an alpha-shape but not every alpha shape is a convex hull. By definition, for each real number α, it is defined as the concept of a *generalized disk of radius* $1/\alpha$ with the following rules: (i) if $\alpha = 0$, it is a closed half-plane; (ii) if $\alpha > 0$, it is closed disk of radius $1/\alpha$; (iii) if $\alpha < 0$, it is the closure of the complement of a disk of radius $-1/\alpha$. *Alpha shapes* are closely related to *alpha complexes*, subcomplexes of the Delaunay triangulation of the point set. In detail, each edge or triangle of the Delaunay triangulation may be associated with a characteristic radius, the radius of the smallest empty circle containing the edge or triangle. For each real number α, the α-complex of the given set of points is the simplicial complex formed by the set of edges and triangles whose radii are at most $1/\alpha$. *In the case of Intervor, Intervor uses the weighted α-shapes and the α-complex in order to generate the best bounding and docking among the atoms of the complex of the two protein interfaces.*

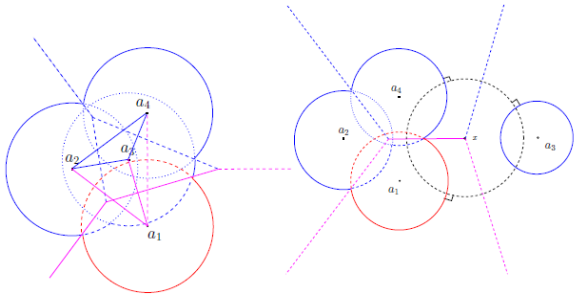

Figure 1 : Figure describing all the above definitions. **Left:** Bicolor interface defined from the α-complex of the atoms. Interface edges are a1a2 and a1a3. **Right:** The sphere centered at x is orthogonal to spheres centered at a1, a3, a4, and is the largest sphere orthogonal to the spheres centered at a1 and a4. Its radius can be used to accept/reject the interface Voronoi facet associated to the Delaunay edge a1a4.

Crystals and structural Water –the AB and ABW models- : As we noticed in a previous section, Intervor only needs a PDB file to operate. So consider a PDB file featuring a complex. For example, for a protein-protein complex, each partner is specified as a collection of polypeptidic chains. All atoms found in the PDB file are either tagged as A or B for the

two partners, W for structural water, X for the remaining atoms. These four tags are called the types of the atoms, and a pair (triple) of atoms is called bicolor (tricolor) if it features two (three) different types. Since we define our interface model from the Voronoi diagram of the atomic balls, there are introduced two models:

- *ABW model:* all atoms found in the PDB file, excepted hydrogens and possibly selected water molecules, are inserted into the Voronoi diagram.
- AB model: all atoms found in the PDB file excepted hydrogens and water molecules are inserted into the Voronoi diagram.

In particular, notice the ABW model is highly relevant for high resolution crystals — resolution better than 2°A, where water molecules are spotted reliably. Notice also that one may wish to retain all such molecules, or only those having a relatively low temperature factor.

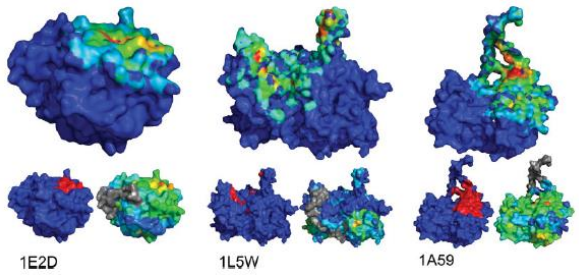

Figure 2: Projection of Voronoi shelling order (large panels), dryness (lower left-hand panel), and conservation (lower right-hand panel) on the molecular surface of homocomplexes 1E2D (left), 1L5W (center), and 1A59 (right); one of the monomers was removed for clarity. Cold (resp. hot) colors represent low (respectively high) values; gray areas denote residues for which conservation information was unavailable.

The AB, AW-BW, ABW interfaces: Any atom involved in at least one interface edge, contributing in complex, is termed an *interface atom*. Pairs selected this way do not account for all pairs within a distance threshold, but it is proved in that such interface atoms form a superset of atoms loosing solvent accessibility. More precisely there exist three types of edges selected interfaces:

- **AB interface:** the interface specified by edges of type AB. This interface describes the contacts between both partners.
- **AW−BW interface:** the interface specified by edges of type AW or BW. This interface describes the contacts between the partners and structural water.
- **ABW interface:** the union of the AB and AW−BW interfaces. This interface positions relatively to one another the previous two interfaces.

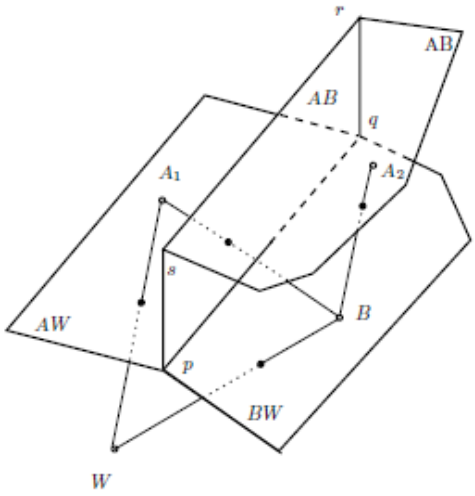

Figure 3: An ABW interface with four atoms A1, A2, B, W: edge pq is shared by three Voronoi facets of types AB, AW, BW and is non manifold; edge rs is shared by two Voronoi facets and is manifold. The boundary net consists of 15 edges (all edges but edge pq).

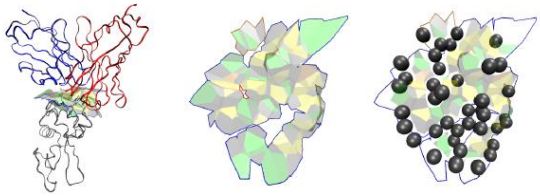

Figure 4: Left: Side view of the 1vfb complex, with chains drawn as ribbons: grey: Lysozyme; blue-red: Fv fragments of mouse monoclonal antibody D1.3. **Center:** Top view of the AB interface in the AB model **Right:** The boundary of the AB interface in the AB model, the AB interface (without boundary) and the structural water molecules in the ABW model.

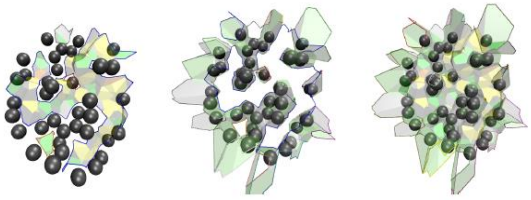

Figure 5: Complex 1vfb, ABW model **Left:** The AB interface and the water molecules **Center:** The AW−BW interface **Right:** The ABW interface. Notice common boundaries have been merged.

Connected Components: The transitive pairwise intersections between atoms. The interface model defines connected components (cc) and significant connected component (scc) as a cc whose surface area is at least a tiny fraction of the AB interface. In the AB model, we report the number of cc and scc, denoted #cc and #scc. For example, a cc of the AB interface corresponds to a collection of A and B atoms such that a restricted atom of type A intersects a restricted atom of type B which intersects a restricted atom of type A and so on. *In the case of Intervor, Intervor uses the Connected Components mostly for statistical purposes in topological issues.*

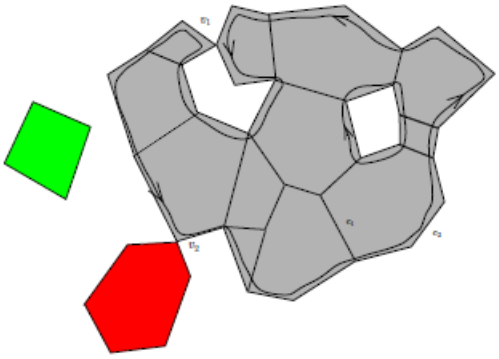

Figure 6: An interface with 3 edge connected components. The red and grey (scc) connected components, connected by a pinched Voronoi vertex, are not edge connected. The boundary of the grey component consists of two loops. Edge e1 is an interior Voronoi edge, edge e2 is a boundary Voronoi edge. CC are also called patches.

Statistics: Intervor can be configured in a large scale and as a result due to the configuration one wants a lot of exported data can be used for statistical studies. In this section, we review the statistics one can compute from the interface model, and mention their relevance to investigate complexes. All information required to compute such statistics are available in the files generated by Intervor.

Further, explanation about the notations of extracted by the Intervor in the output files are provided in the (Cazals, 2006, págs. 26-30).

Experiments

Now that there has been referenced in brief all the related bibliography, and anything one needs to take under consideration so as to understand the results of our subsequent experiments, we provide in this section all the requisites for the conduct.

- **Dataset:** Taking under consideration the importance of the dataset gathering, as it always consists the cornerstone in every experiment, we abutted that the groups upon which we want to reveal a hidden difference are most preferred to have functional correlations among their sub-units and that it is more preferred not to group the proteins according to their structural correlations. As a result, we concluded in distinguishing the dataset in groups of families of proteins according to their molecular function, and especially the families we took for our experiments are: proteins involved in (i) transducer protein activities, (ii) transporter protein activities, (iii) receptor protein activities, (iv) protein binding transcription factor activities, (v) enzyme regulator activities.
  What is more, a few more criteria were selected in order to eliminate from our dataset proteins that are not the appropriate for use with the Intervor. These criteria are made up with (i) a resolution better than 2Å, and especially we selected a resolution between 2Å-3.5Å, so that water molecules can be shown and those would be only the significant water molecules that can participate in connections on the Intervor interface due to their measures, (ii) a min number of oligomeric state set to 2, because it is very possible to exist complexes found by Intervor in proteins consisted of oligomers, (iii) a homolog removal of 95%, so that the statistics that we will conduct should be more trustworthy, as most of the proteins are not homologs. As a fourth criterion (iv) we selected our proteins to have free ligands, because it is possible for the ligands to participate in Intervor

interface. This criterion may have also some disadvantages in other statistics as for the Intervor it is not important for ligands to exist. Nevertheless, we decided to include this selection in order to find a statistic for ligands especially, that may involve in the interface. Finally, in order for the Intervor to operate we selected (v) that the chain is, as it is reasonable, of a protein type. The queries are provided in a separate file with the homonym filename, "Intervor Project\experimental datasets and notes\queries.txt".

- **Intervor I/O:** Having created a script that downloaded all the pdb files of the above groups of proteins we gave these proteins as an input in Intervor, which we called in the same script with a system call and stored each pdb's resulted log output files in folders named after the pdb file name. The Perl script is named as "Intervor_Parser.pl" (and "Intervor_Parser_Bench.pl") found in the main directory.
- **Intervor Log Conventions:** Intervor generates files containing pieces of information about interface atoms and their contacts. When dumping a given atom, it is used the plain PDB format, or the reduced PDB format, which consists of the following four-tuple: (i) atom serial number, (ii) PDB atom name, (iii) residue name and (iv) chain id.

  All atoms of a PDB file are assigned to five types: (i) A and (ii) B for the two partners, (iii) Wb and (iv) Wi for the bulk and interface water molecules, and (v) X for the remaining atoms (It is worthy to mention here that water molecules loaded into the Voronoi diagram are classified either as interfacial or bulk). In the files generated by Intervor, these types are respectively denoted (i) IA A, (ii) IA B, (iii) IA Wb, (iv) IA Wi, (v) IA X (-the Prefix IA stands for Interface Atom-).

  The five atomic types are used to define types for pairs, in particular types AB, AWi and BWi which are used to define the AB and AW-BW interfaces. In the output files, these types are denoted IE AB, IE AWi and IE BWi (-the prefix IE stands for Interface Edge-).

  These five types are also used to define so-called manifold interfaces: a manifold interface is a subset of the Voronoi diagram such that any Voronoi edge is incident to at most two Voronoi faces. Of particular interest are the following manifold interfaces, also called bicolor interfaces: AB;AWi;BWi. The collection of all tiles of type AWi and BWi define the AW-BW interface. In the output files, these interfaces are respectively denoted MI AB; MI AWi; BWi; MI AWi BWi (-the prefix MI stands for Manifold Interface-). It is noted in passing that the ABW interface is not a manifold one, since three Voronoi facets can be incident on the same Voronoi edge.
- **Logs and information included:** Intervor generates six files. These files are self-contained, whence some redundancy, and provide a description of the two complementary aspects of interface modeling: interface atoms and Voronoi interface. These files are the following ones:
  - *Pdb_name_log-IV.txt:* general. This file is only generated when Intervor is called with the ***--log*** option; otherwise, its content is dumped onto the standard output. Information provided is high-level information: number of interface patches and corresponding statistics, number of atoms, number of pairs, etc. The remaining _les detail this information.
  - *Pdb_name_iar-IV.txt:* interface atoms. The list of interface atoms in text format, together with selected annotations.
  - *Pdb_name_iar-IV.pdb:* interface atoms. The list of interface atoms in PDB format. The atom serial number reported is that of the original file.
  - *Pdb_name_contacts-IV.txt:* pairs of atoms. The pairwise contacts.

- *Pdb_name_interface-IV.txt:* Voronoi interface. Geometric and topological description of the Voronoi interface.
- *Pdb_name_patches-IV.txt:* patches of the Voronoi interface. Dissection of the interface atoms as a function of the patches they participate to.

- **Information used for statistics:** As described above, we decided to take all the information we wanted to make our statistics from the *"pdb_name_log-IV.txt"* files. Information provided in this file splits into three sections. The first two are present in any case; the third one is found if and only if interface water molecules are detected. A bigger analysis in the included descriptive units can be found in (Loriot, 2009). The work our team has done was to create another Perl parser, *"Log_Parser.pl"* which finds the sum, mean values and standard deviations of areas in the interface such as: number atoms of A,B,Wb,Wi,X in the pdb, number atoms of A,B,Wb,Wi,X in the pdb interacting in the interface, atoms of certain types as defined in the bibliography ("std-atomic-radii.txt") and presented in a following section, number of connected components, number of involved chains, etc.
- **Statistic Results:** In order to choose the appropriate statistics to proceed with our analysis, we studied the log files description (Loriot, 2009) and several studies (Boovier, 2009), (Cazals, 2010) that have analyzed protein families in a same fashion. The statistics we decided to gather from the log files, are candidate features that might be used to discriminate the protein families that are included in our dataset. In fact, we chose the following statistics – features: As structural features we considered the (1)Triangles and (2)Tetrahedra that arose by the construction of the Delaunay triangulation of each protein interface, the number of atoms of each atom type – Partner, (3)Partner A, (4)Partner B, (5)Partner Wb, (6)Partner Wi, (7)Partner X, the equivalent atoms that participate in the interface, (8) interface atoms A,(9) interface atoms B,(10) interface atoms Wb,(11) interface atoms Wi,(12) interface atoms X, the chemical annotations of the interface atoms, (13-27) Cali, Caro, Cpep, Nhdb, Naro, Nchp, Ohbd, Opep, Ochm, Owat, Sh, Pdna, Opd, Orib, Unk, the number of connected components, (28) cc AB and (29) cc AB_AW_BW, the number of interface edges (30) AB, (31) AWi and (32) BWi and finally the chemical annotation of the surface area, (33-42). As a summary of the proteins that are included in each protein group, we provide the number of the proteins that contained water molecules in their interface: For the protein group with the (i) molecular transducer activity proteins, from the total number of 70 proteins, the 17 included water molecules in their surface, for the protein group of the (ii) transporter activity proteins, from the total number of 30 proteins, the 7 included water molecules in their surface, for the protein group of the (iii) receptor activity proteins, from the total number of 31 proteins, the 9 included water molecules in their surface, for the protein group of the (iv) protein binding transcription factors proteins, from the total number of 38 proteins, the 11 included water molecules in their surface and for the protein group of the (v) enzyme regulator proteins, from the total number of 191 proteins, the 56 included water molecules in their surface. In the next section, statistic plots will be provided for each of the features described above. The histograms will depict the man values of each feature for each protein group, normalized by the respective standard deviation.
- **Statistic Plots:** In the following diagrams we try to spot the differences between the protein Groups for the several features described above, in order to be able to discriminate them. Our conclusion will be summarized in the last section of our project analysis.

  We have created 42 histograms, each one representing a feature, and we have computed

the normalized value of each feature for the respective protein Groups. The protein Groups that we took into consideration are proteins involved in (i) transducer protein activities, (ii) transporter protein activities, (iii) receptor protein activities, (iv) protein binding transcription factor activities, (v) enzyme regulator activities. For each of the five protein Groups, we have created two sub-Groups. The first sub-Group of each Group, contains proteins from several species, unlike the second sub-Group that contains only the human proteins of the respective protein family. So, there are ten Groups to take into consideration, Group 1 (molecular transducer activity - all organisms), Group 2 (molecular transducer activity - human), Group 3 (transporter activity - all organisms), Group 4 (transporter activity - human), Group 5 (receptor activity - all organisms), Group 6 (receptor activity - human), Group 7 (protein binding trancription factor - all organisms), Group 8 (protein binding trancription factor - human), Group 9 (enzyme regulator activity - all organisms), Group 10 (enzyme regulator activity - human).

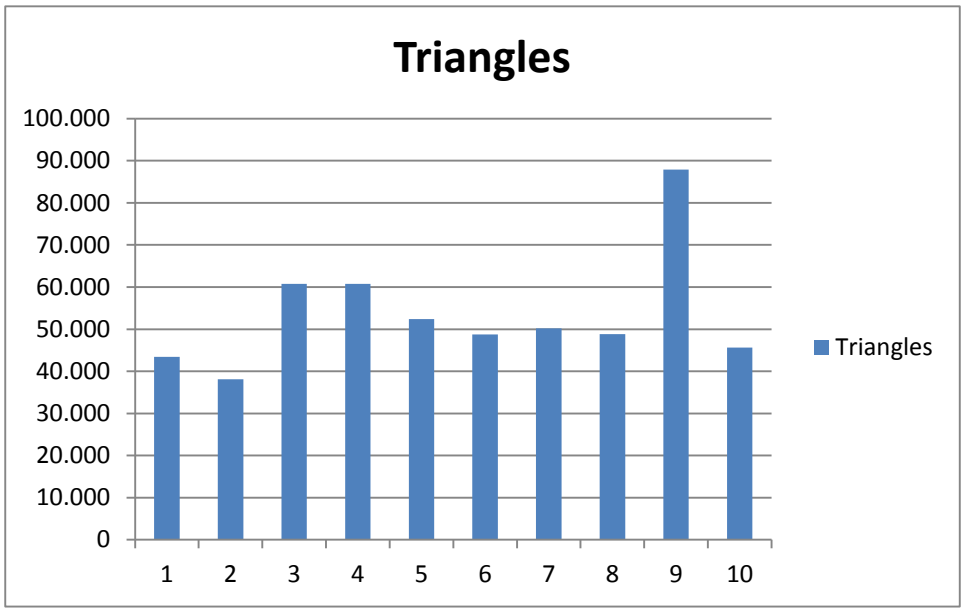

Diagram 1

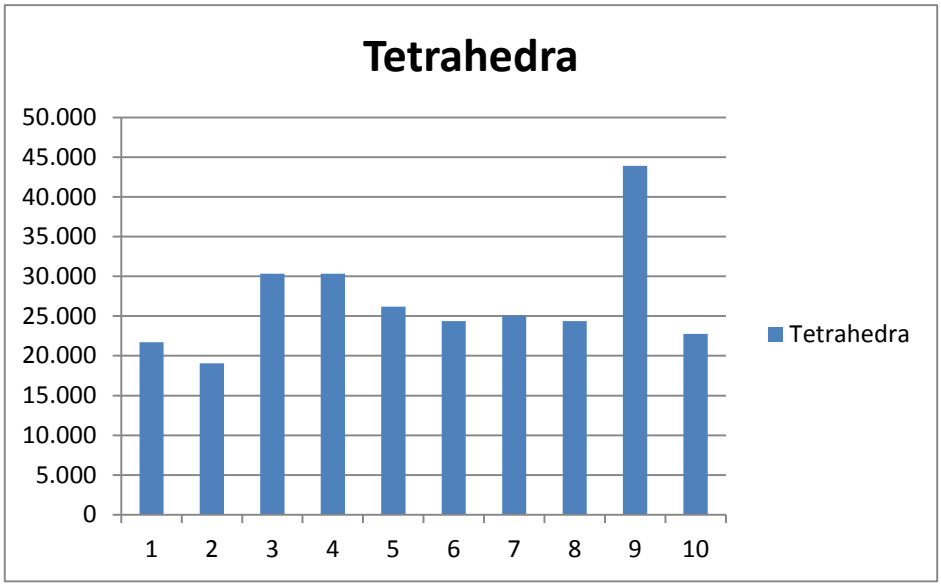

Diagram 2

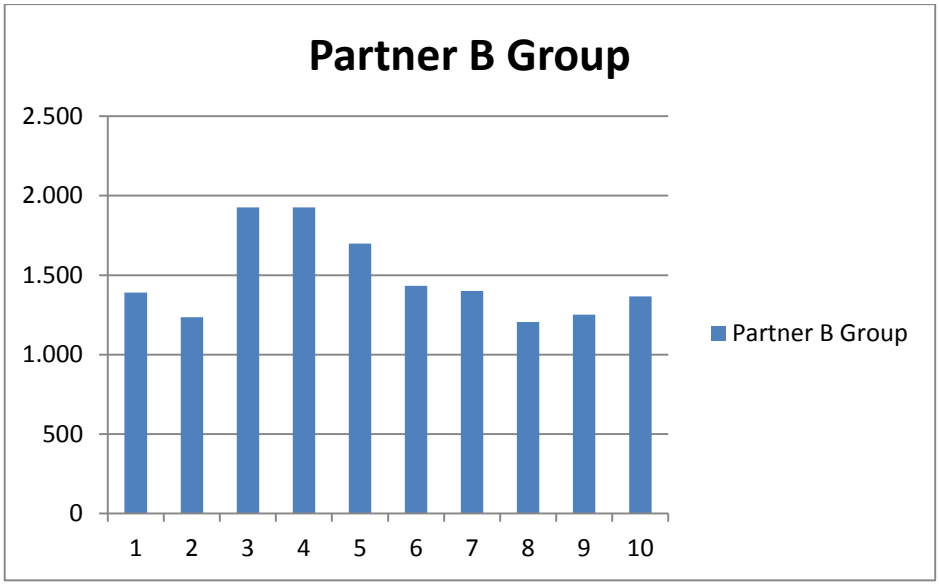

Diagram 3

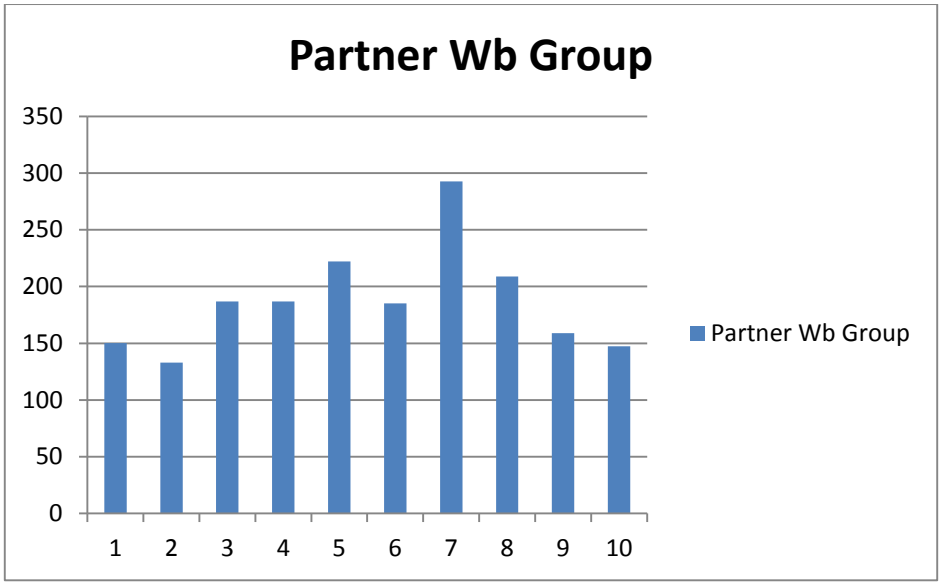

Diagram 4

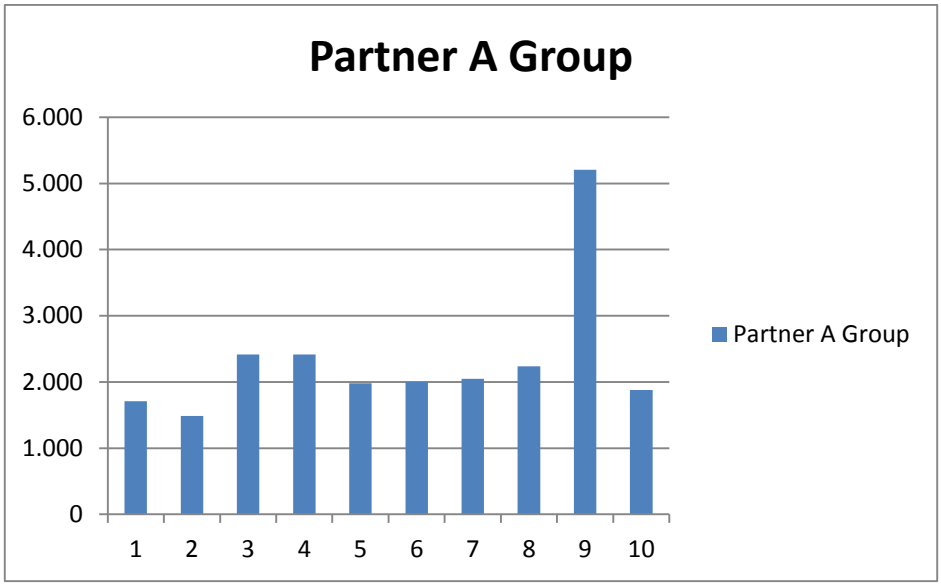

Diagram 5

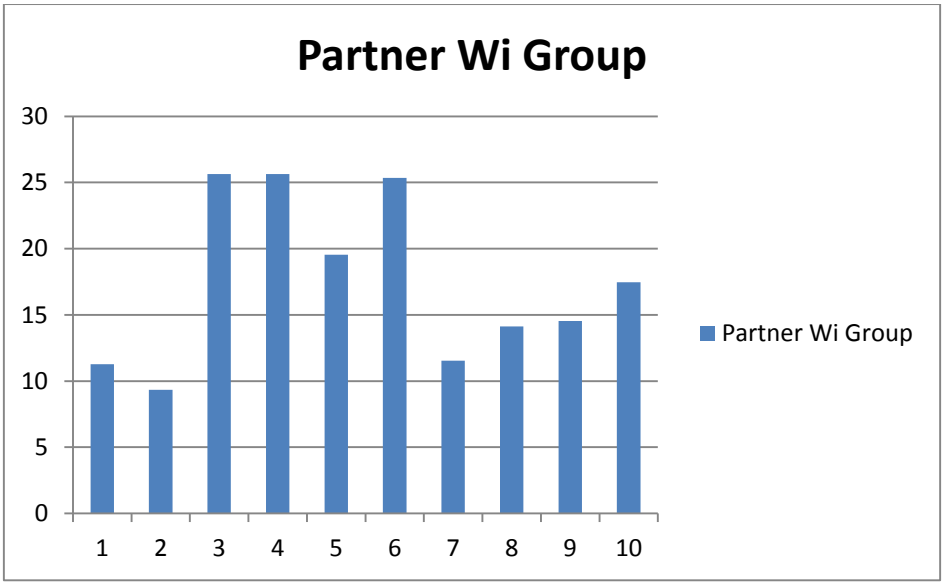

Diagram 6

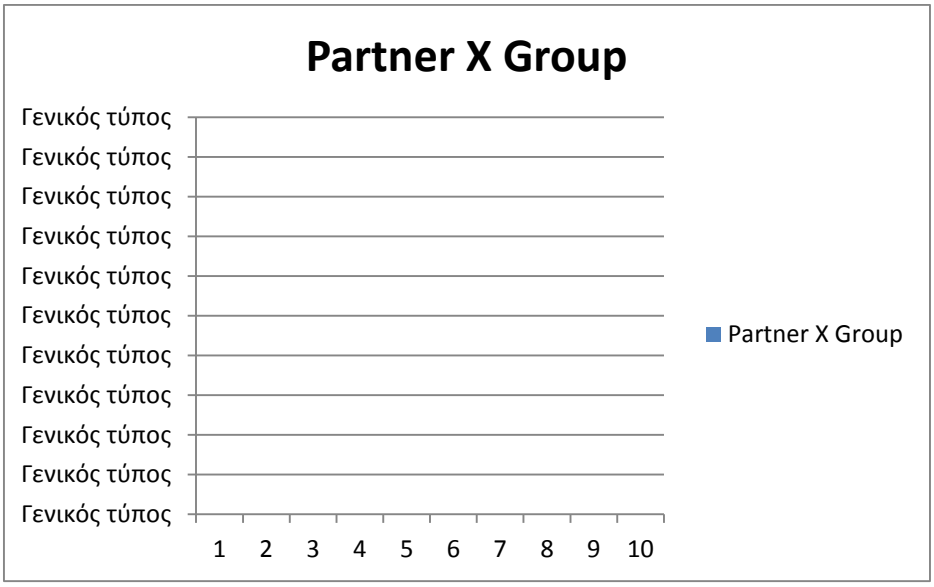

Diagram 7

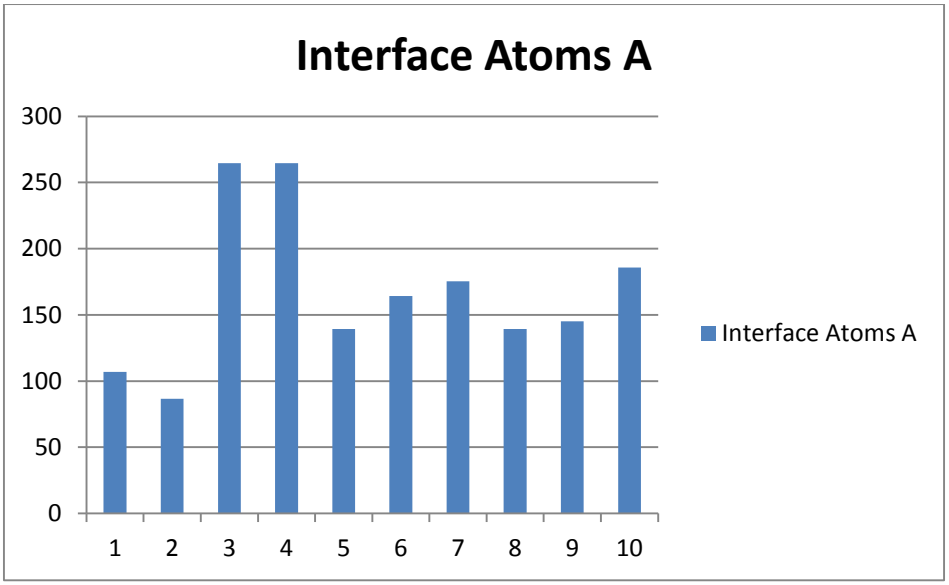

Diagram 8

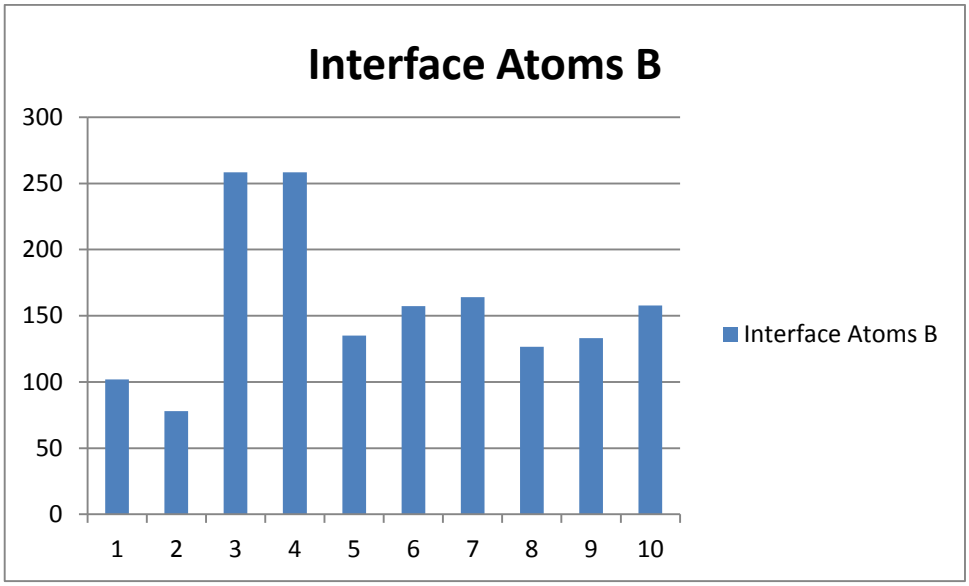

Diagram 9

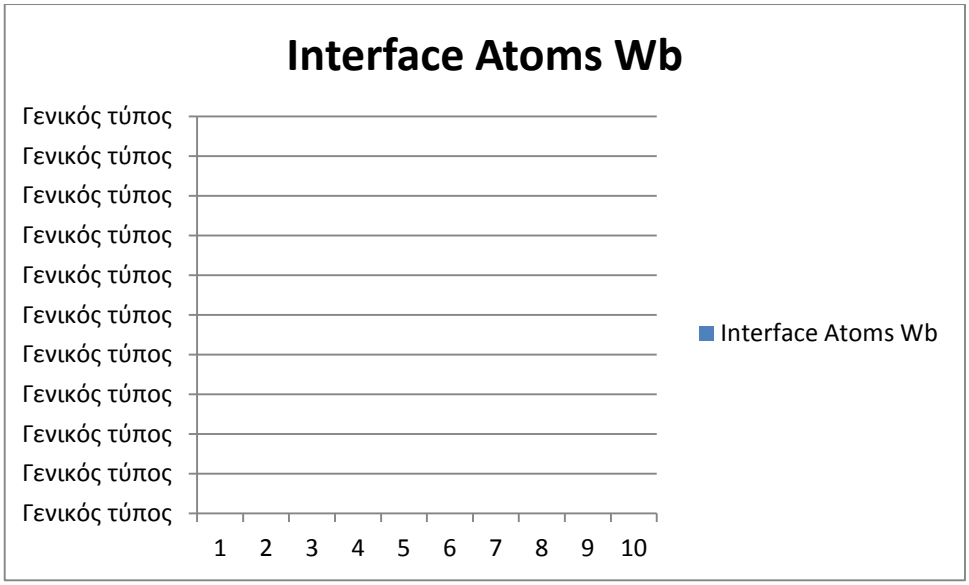

Diagram 10

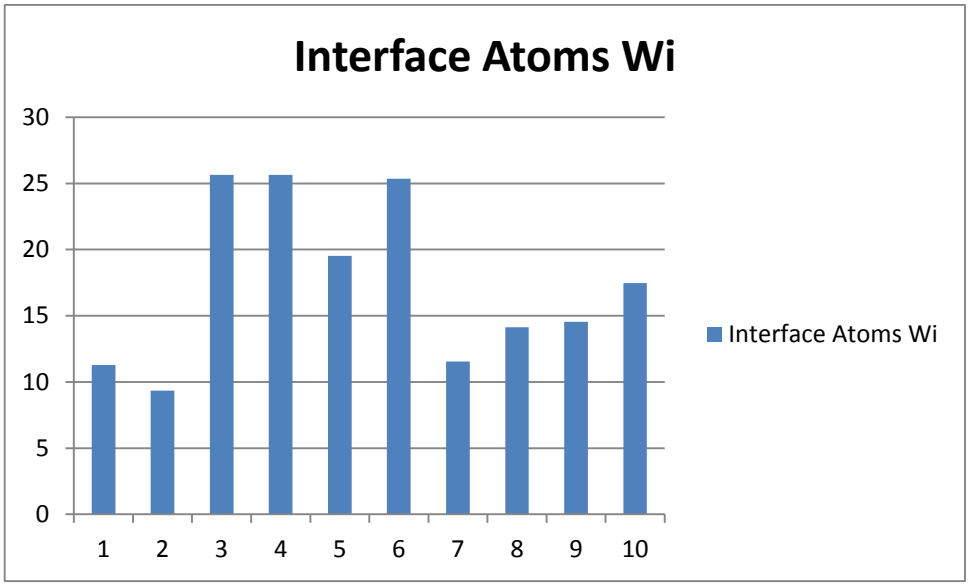

Diagram 11

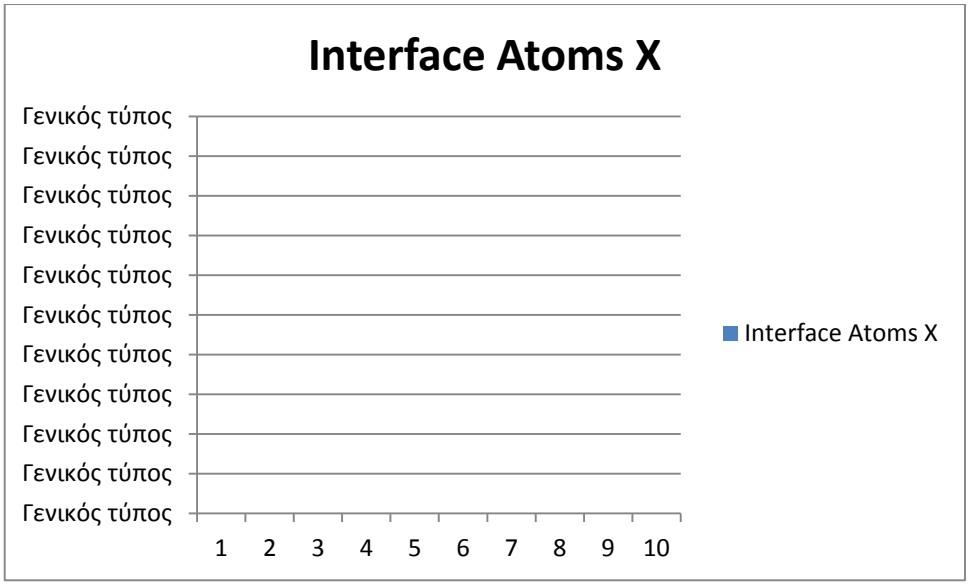

Diagram 12

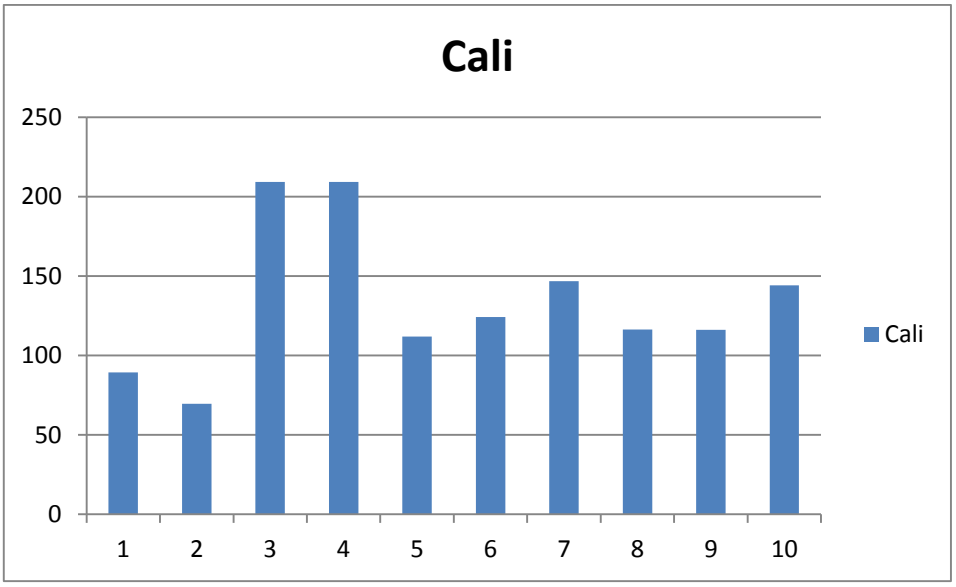

Diagram 13

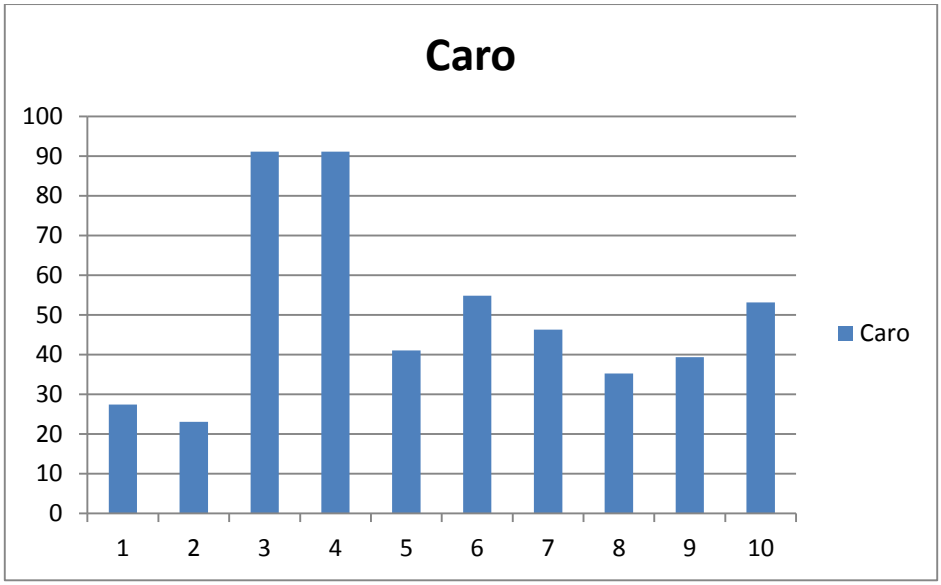

Diagram 14

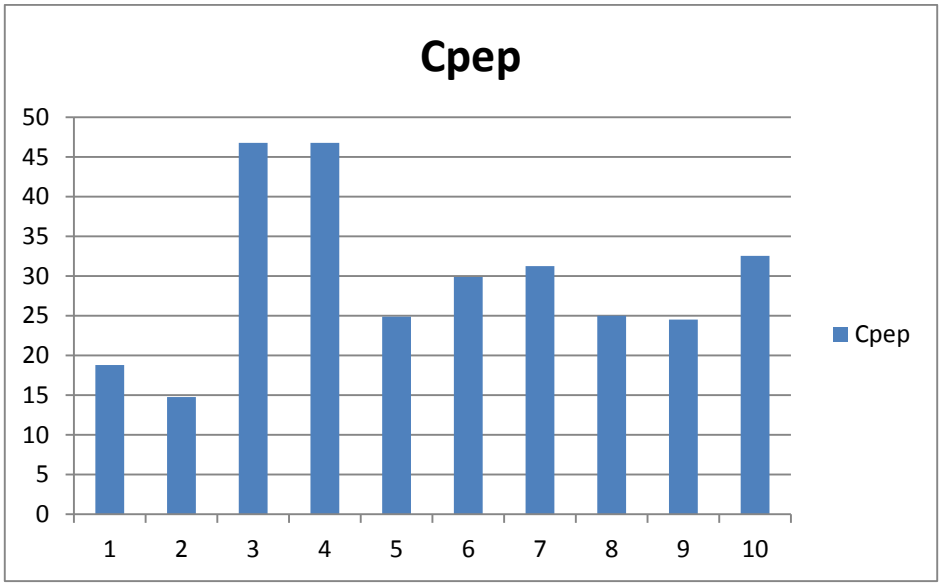

Diagram 15

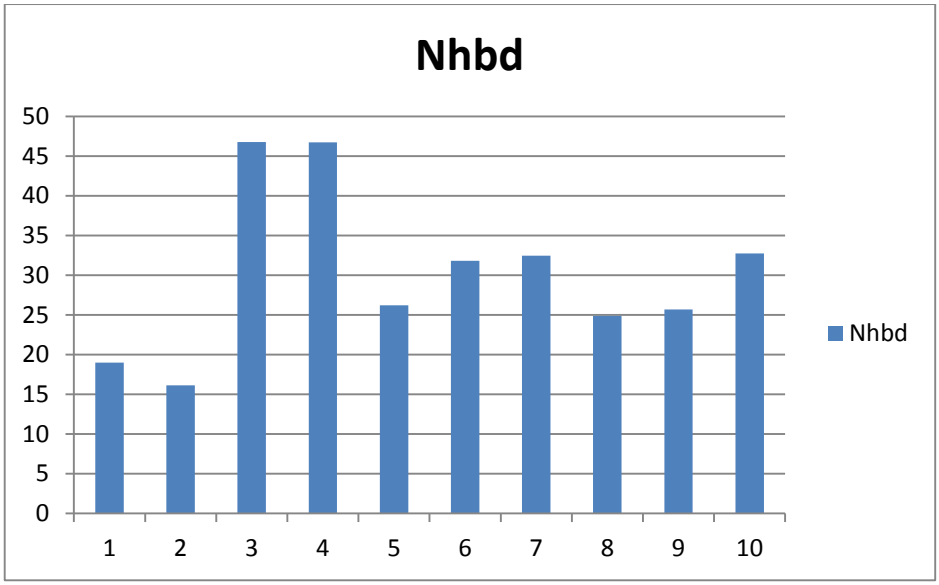

Diagram 16

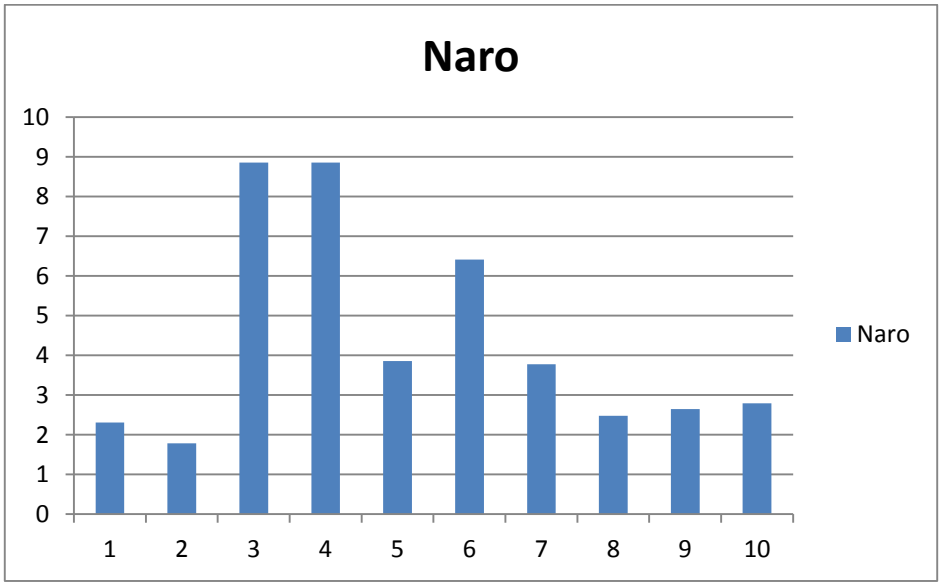

Diagram 17

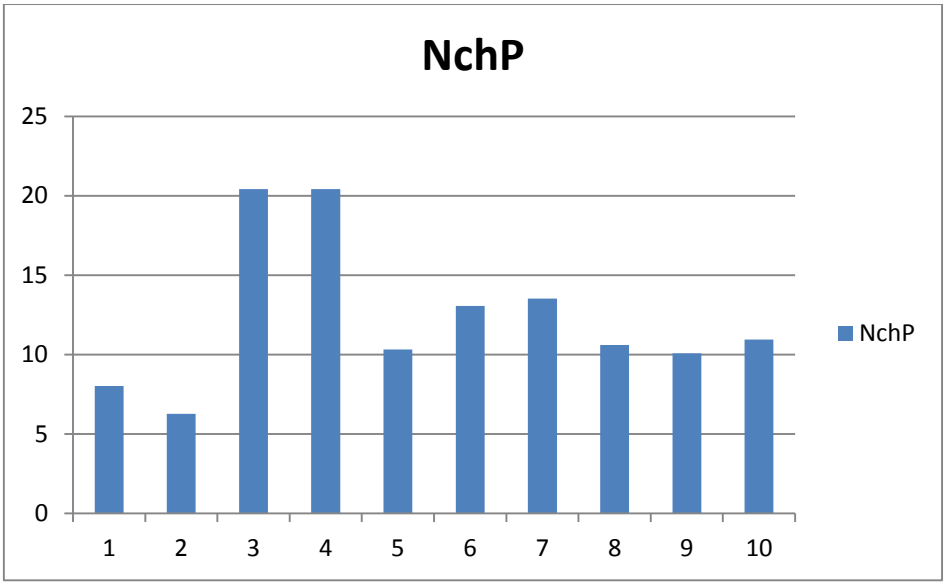

Diagram 18

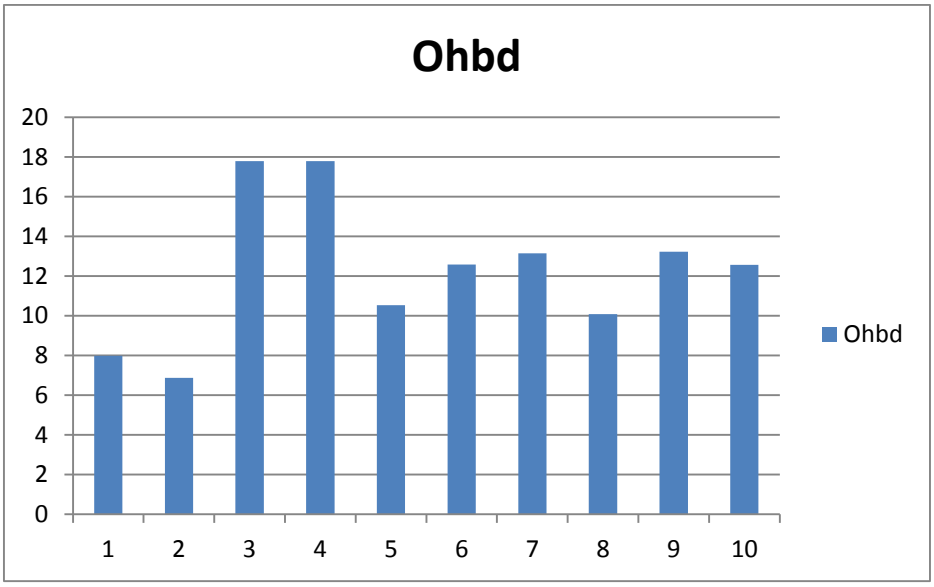

Diagram 19

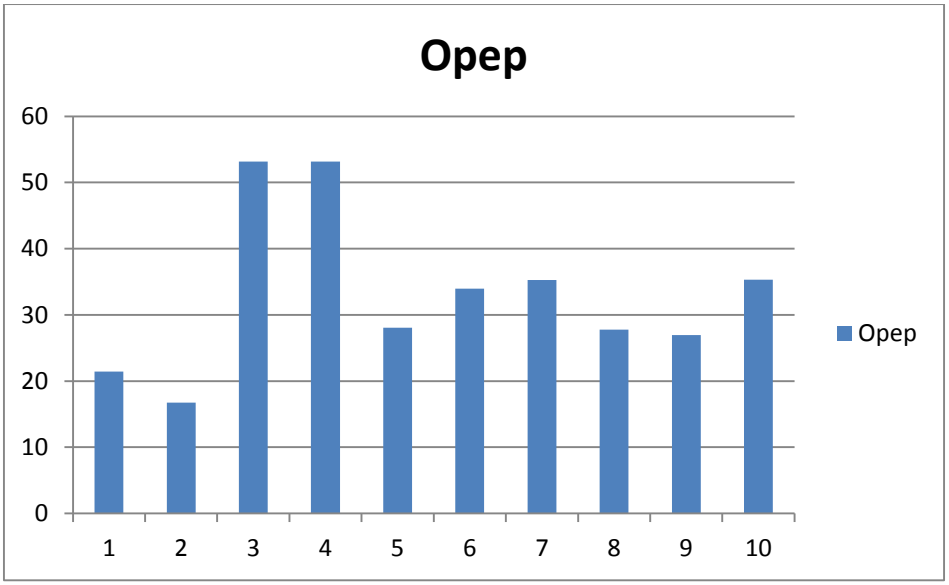

Diagram 20

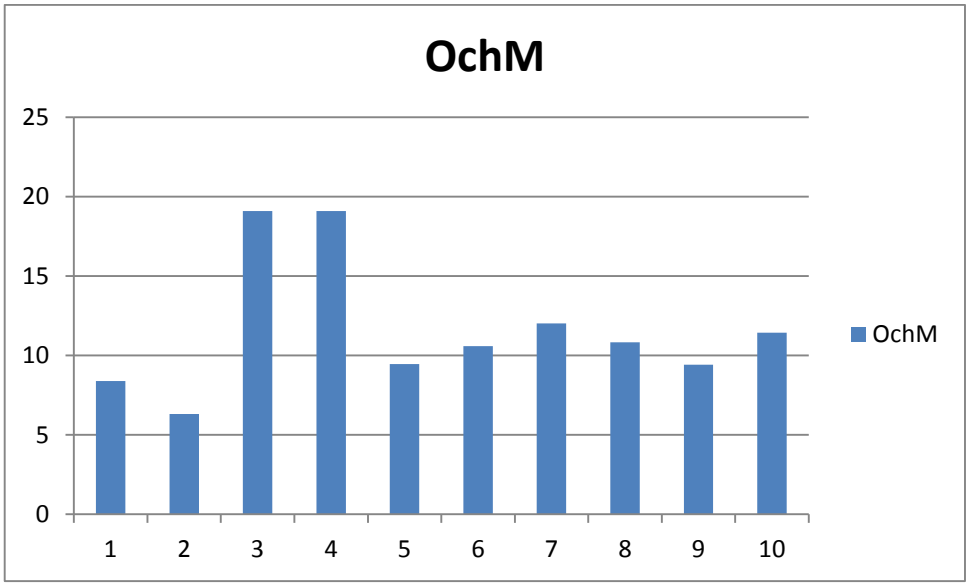

Diagram 21

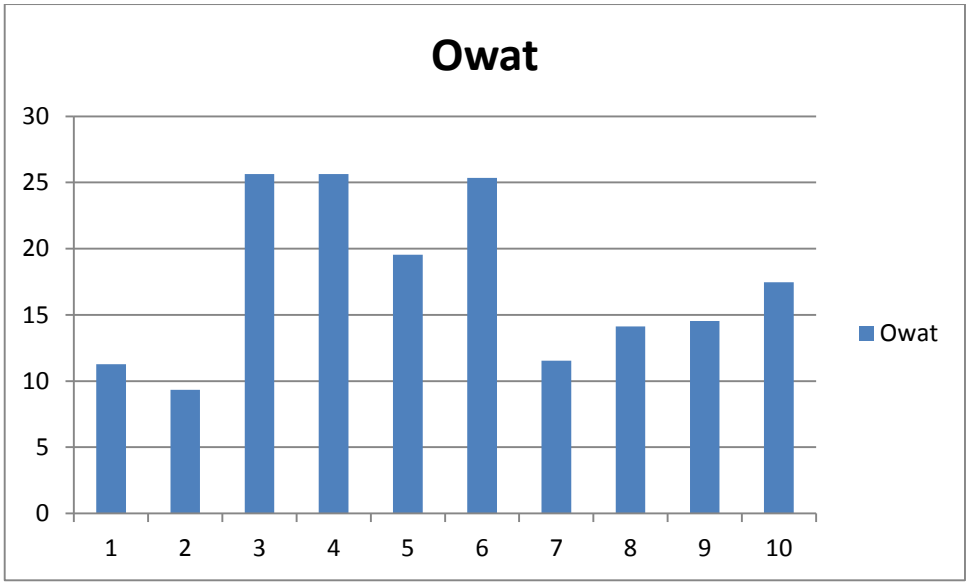

Diagram 22

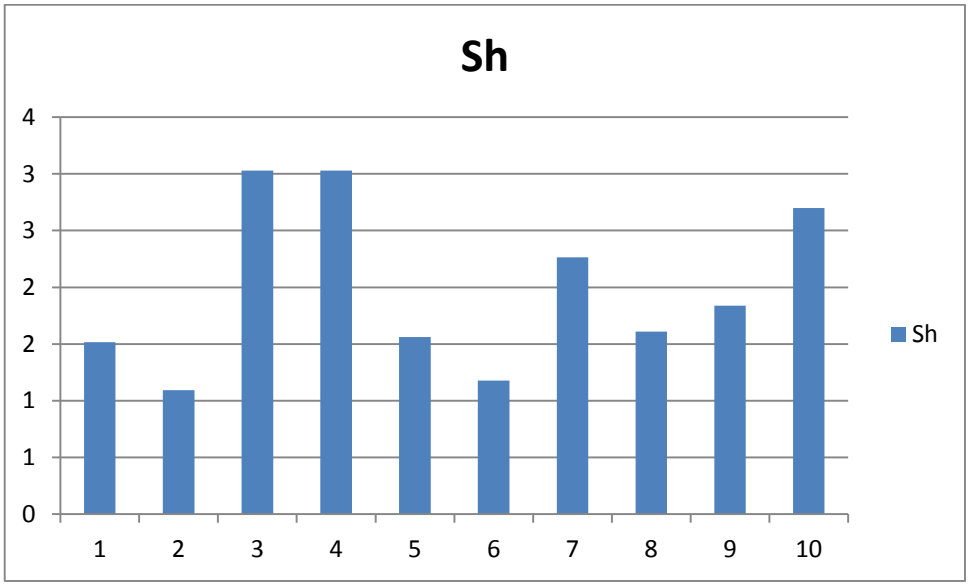

Diagram 23

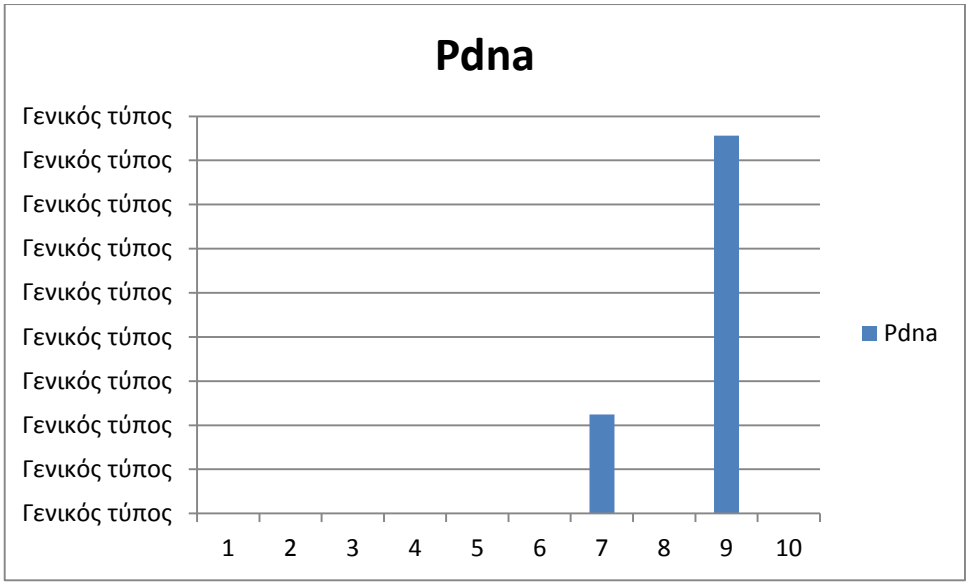

Diagram 24

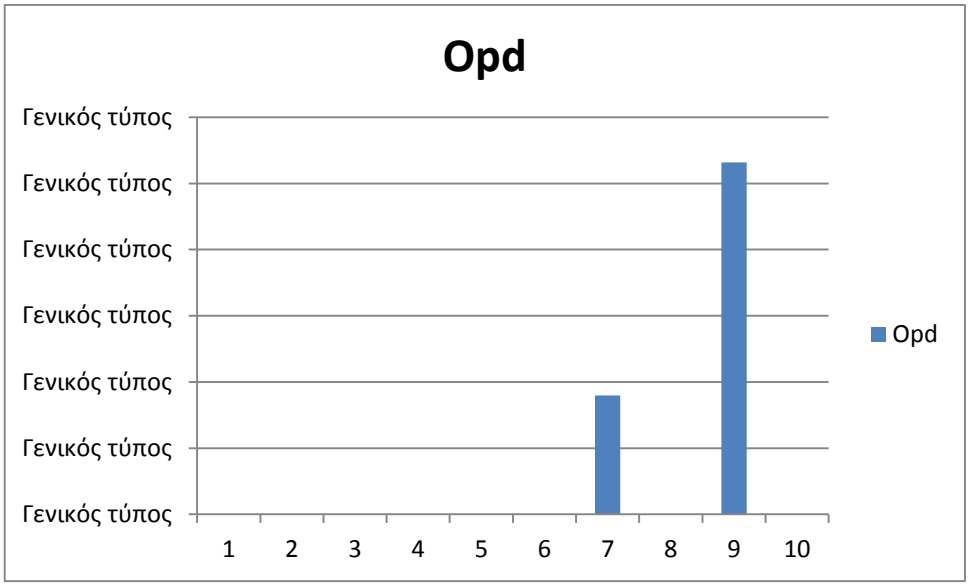

Diagram 25

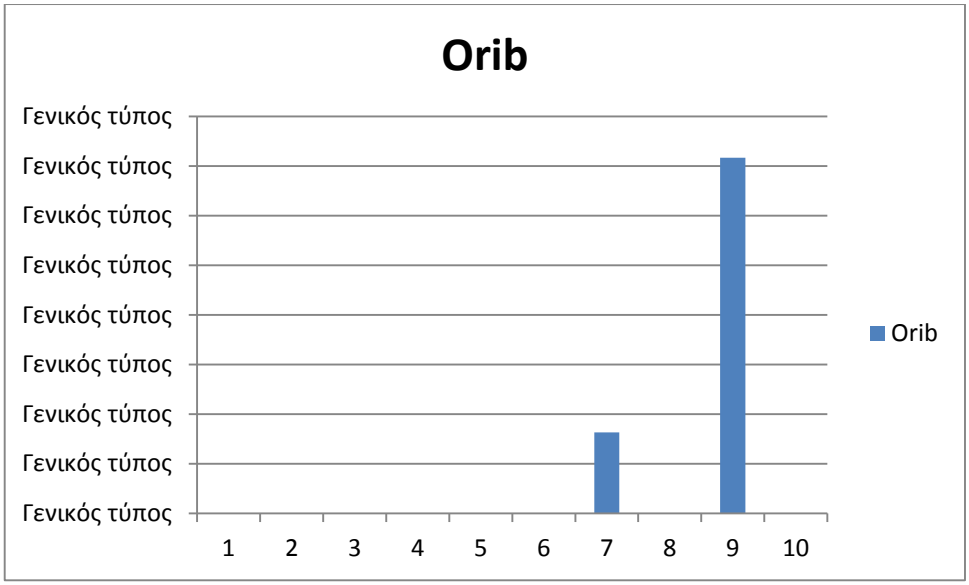

Diagram 26

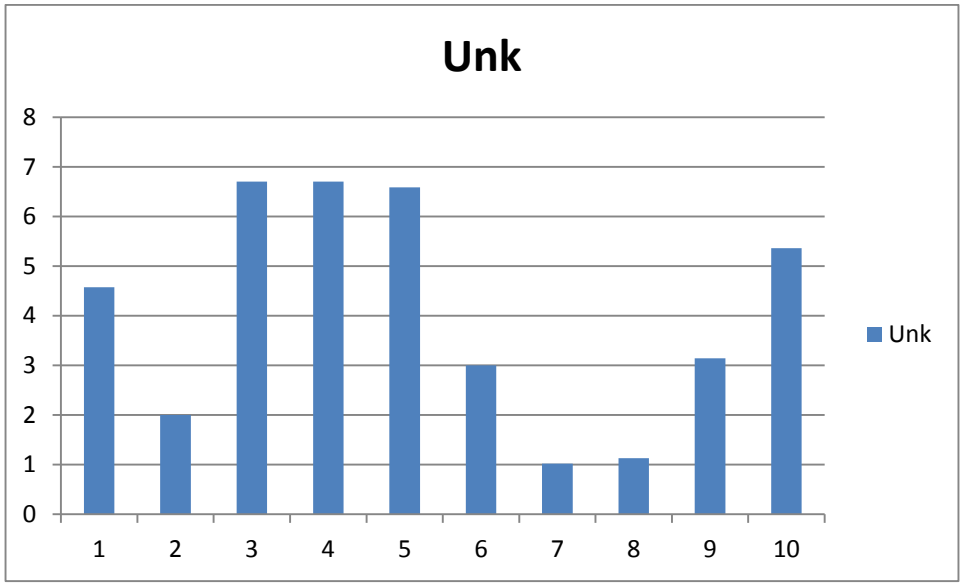

Diagram 27

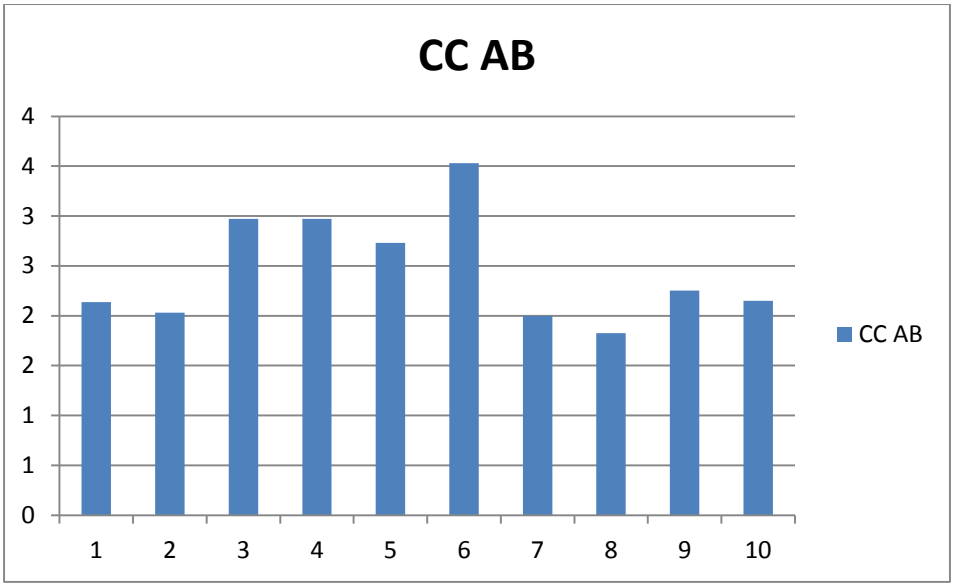

Diagram 28

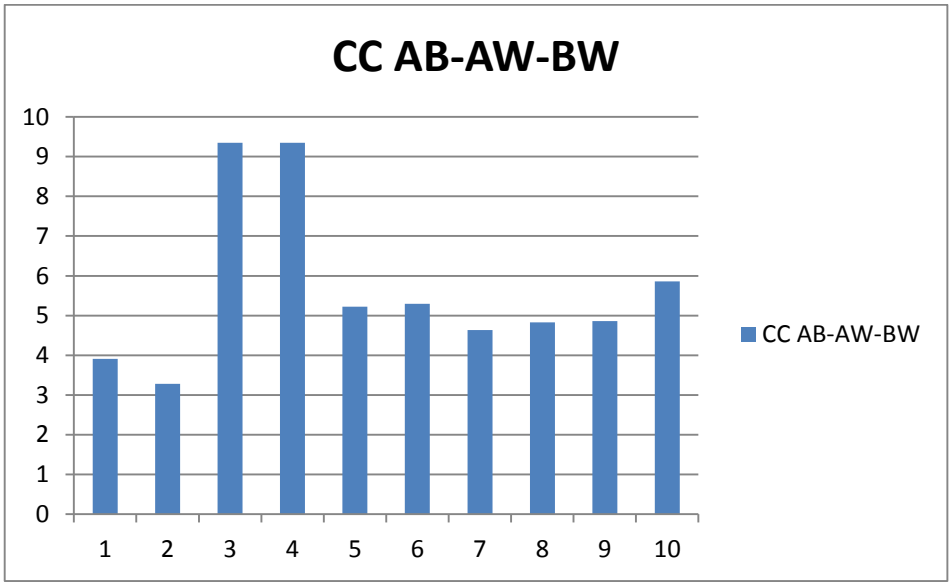

Diagram 29

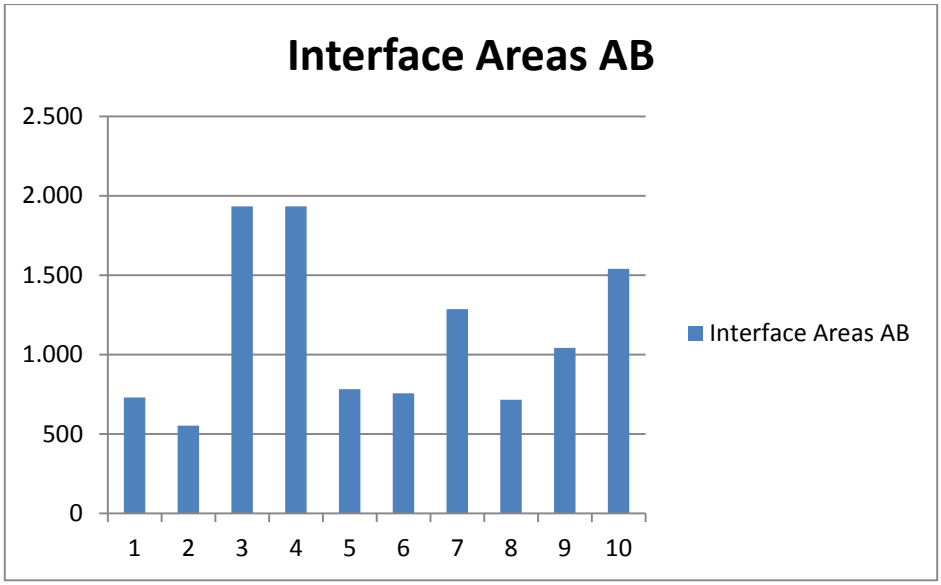

Diagram 30

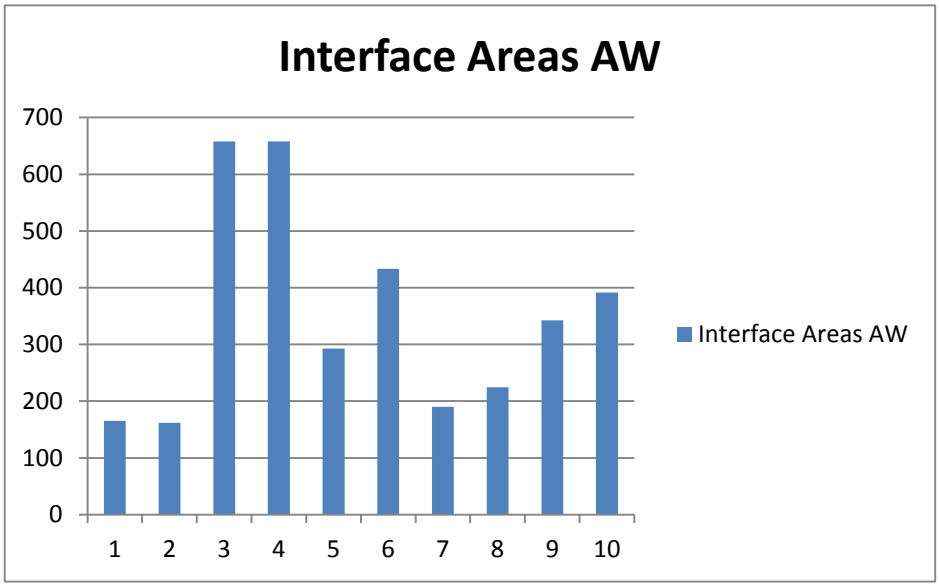

Diagram 31

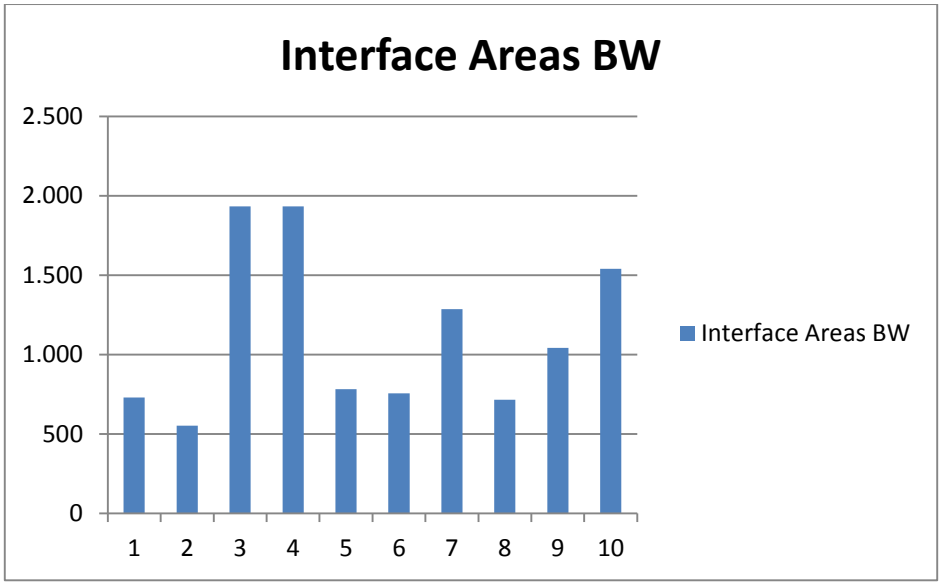

Diagram 32

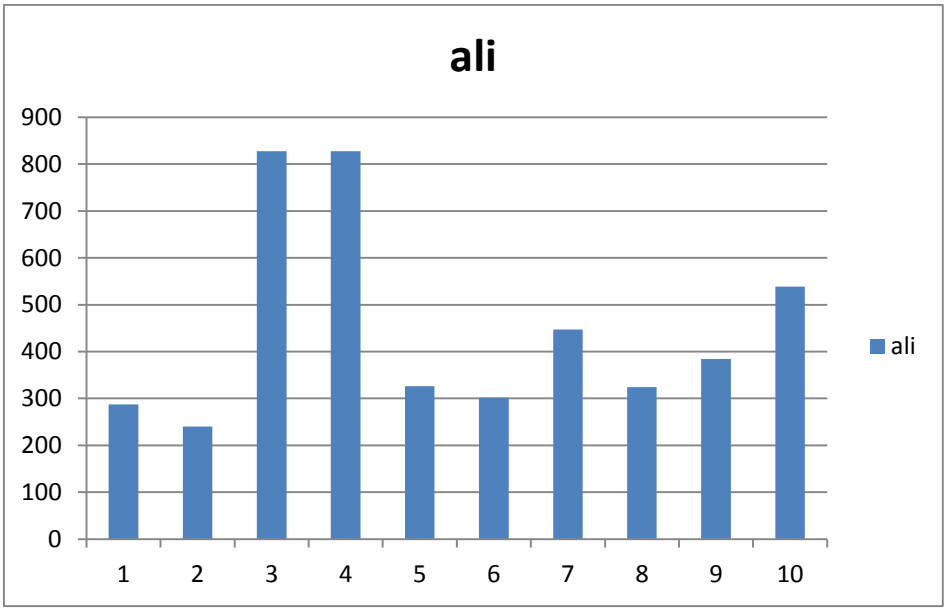

Diagram 33

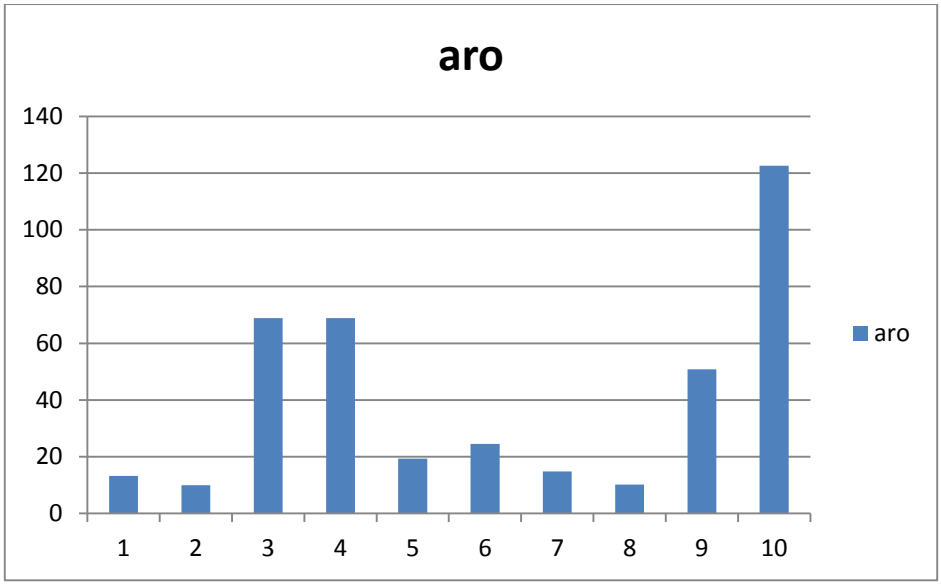

Diagram 34

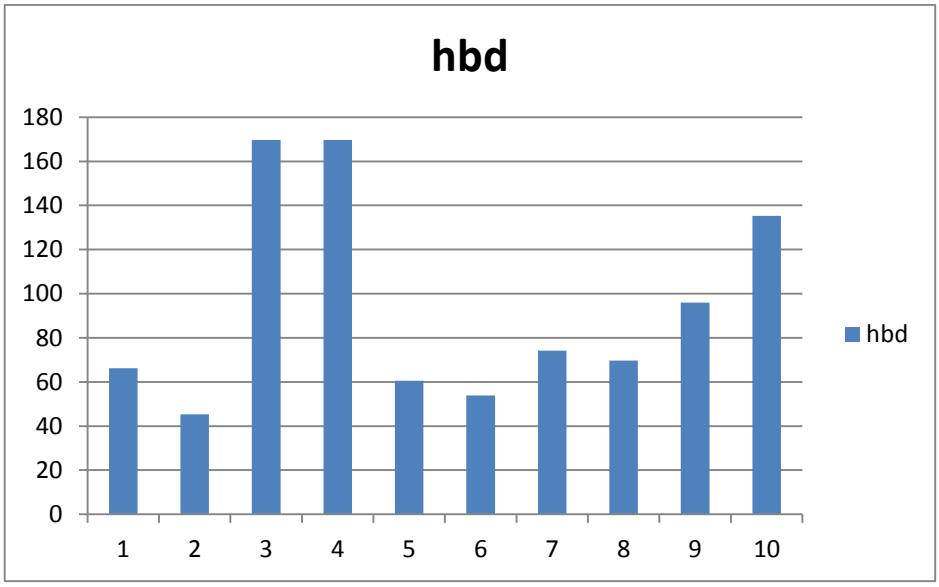

Diagram 35

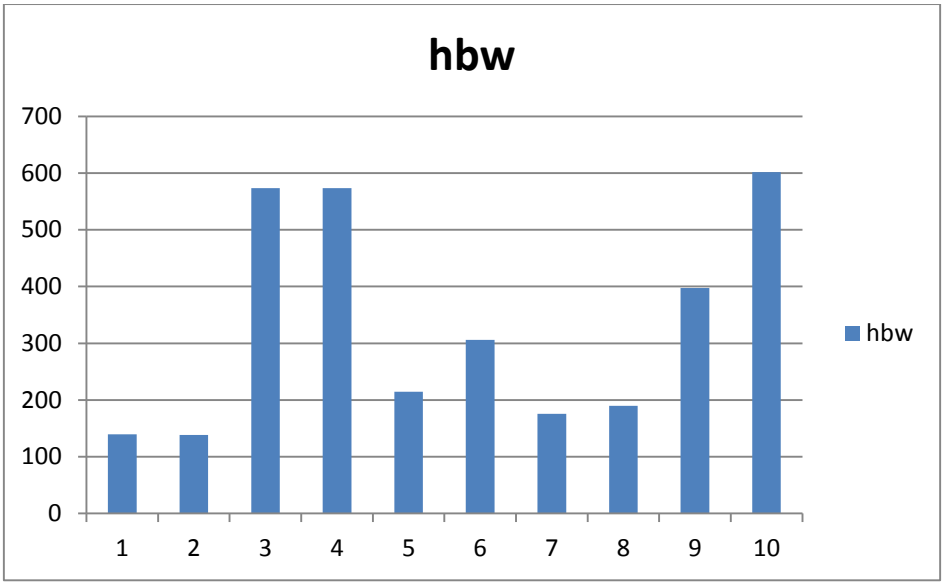

Diagram 36

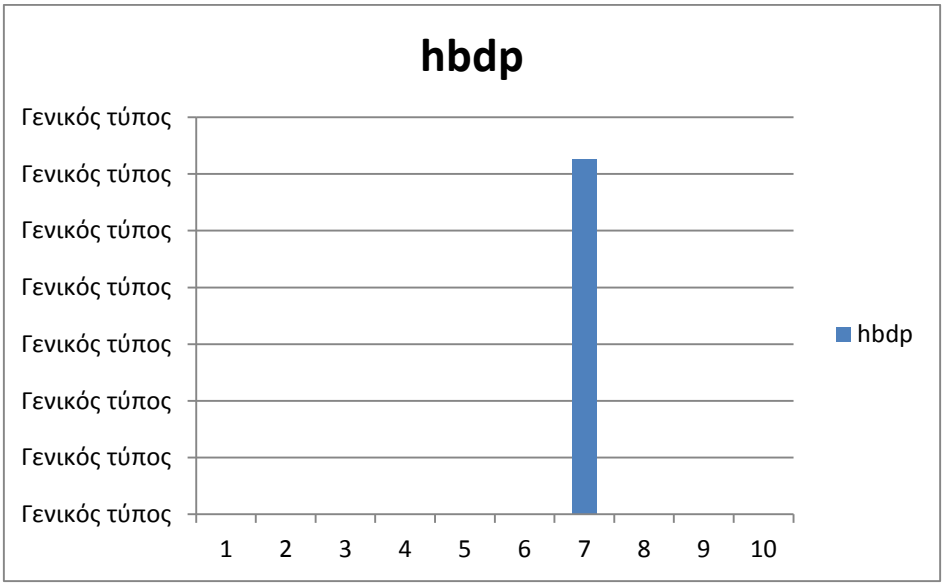

Diagram 37

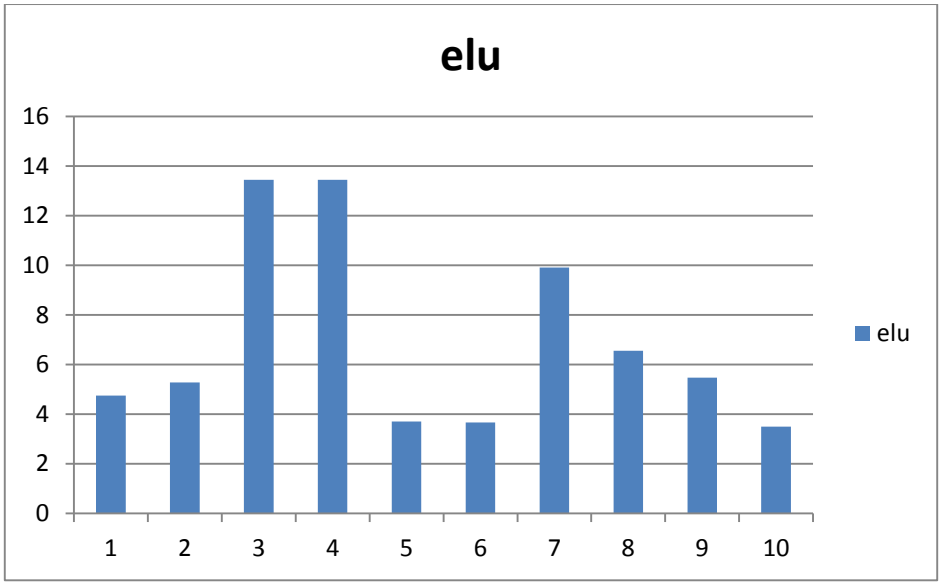

Diagram 38

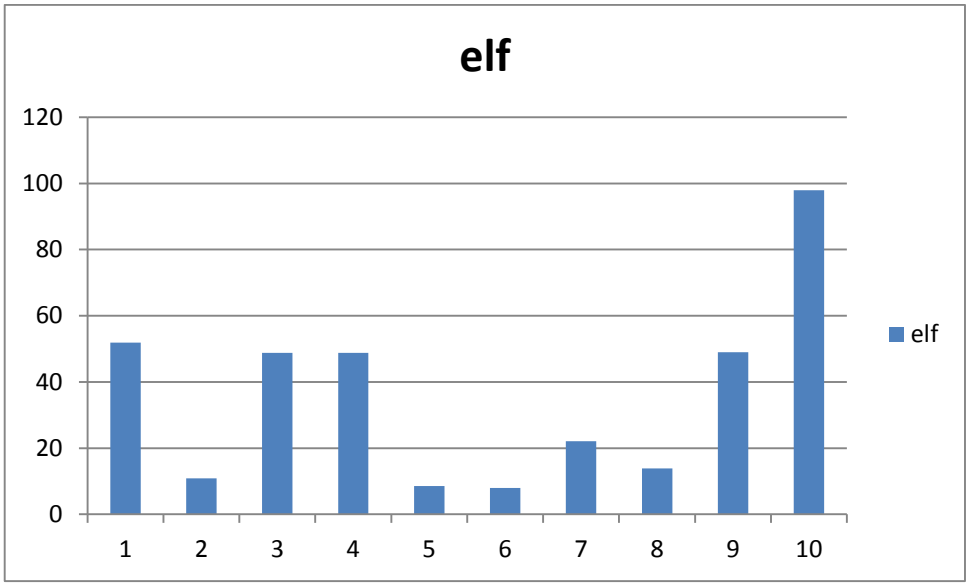

Diagram 39

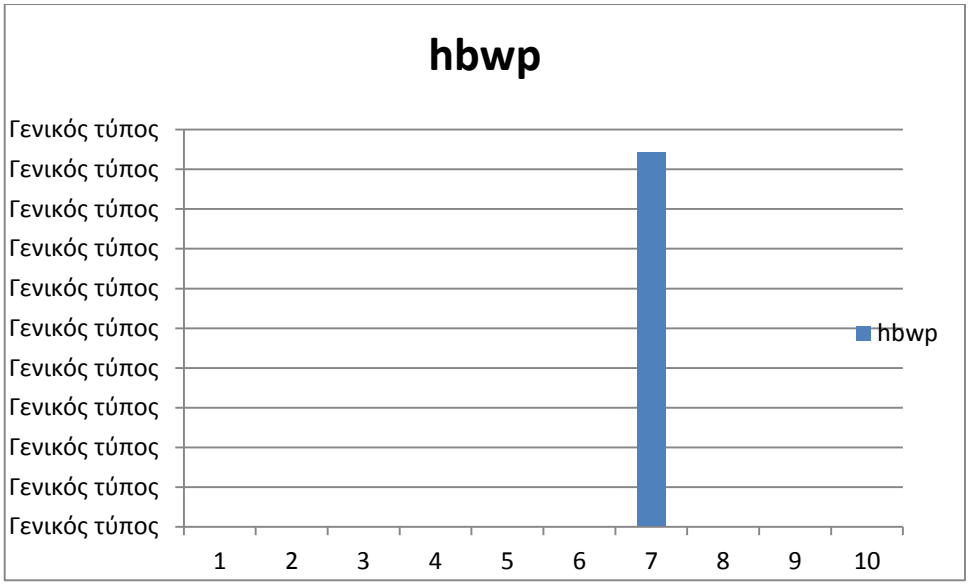

Diagram 40

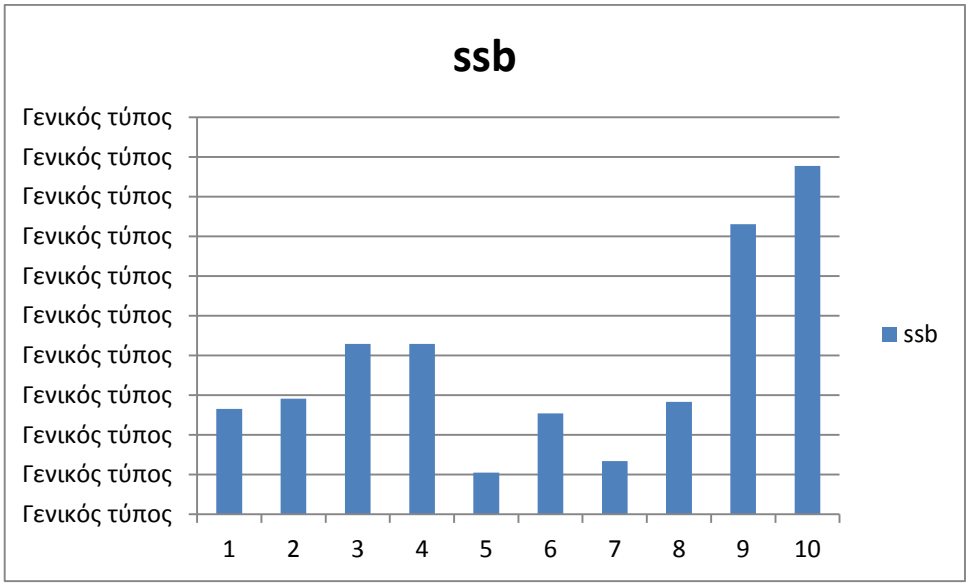

Diagram 41

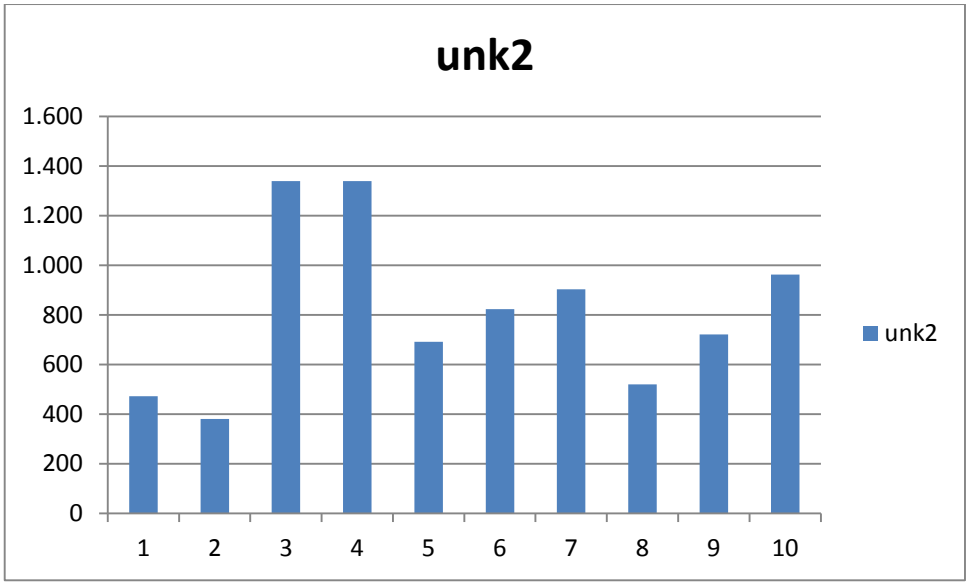

Diagram 42

Conclusions: According to the values of the features we have used for our analysis, the discrimination of the protein Groups of the dataset can be achieved in two different approaches. In the first approach we tried to discriminate the pairs of the protein Groups that belong to the same protein family, in order to identify the classification power of some individual features. The first member of each pair contains all the proteins of the protein family, unlike the second pair that includes only the human proteins of the protein family. For the Group 1 and Group 2 that belong to the molecular transductor activity protein family, the features (27),(35) and (39) seem to be the appropriate features to discriminate the two Groups, with greater values for the Group 1. These features are connected with the chemical annotations of the interface atoms and the chemical annotations of the surface area as described above. Group 5 and Group 6 are discriminated by the features (17),(27) and (41) that are also connected with the chemical annotations of the interface atoms and the chemical annotations of the surface area. Feature (27) has greater values for Group 5, while features (17) and (41) take larger values in the proteins of Group 6. Group 7 and Group 8 are discriminated by the features (5),(17),(23),(30),(38) and (42). The feature (5) is connected with the number of the water molecules that are present in the proteins, but not necessarily in the interface atoms, so it is not a good indicator for Group classification. Besides, it is larger for the proteins of Group 7. The feature (30) represents the number of the connected components of the interface Partners A and B, that is much bigger for Group 7. The features (17) and (23) are connected with the chemical annotations of the interface atoms, while (38) and (42) are connected with the chemical annotations of the surface area. All of these features have greater values for the Group 7. Finally, for Group 9 and 10, we observe that the features (1),(2),(3),(13),(17), (23),(27) ,(32) and (33) have large difference in their values. Specifically, features (1-3) have greater values in the proteins of Group 9 and may indicate some difference in the structural characteristics of the proteins of the two Groups. Features (13),(17) and (23) are connected with the chemical annotations of the interface atoms, while (33) is connected with the chemical annotations of the surface area. All these features that refer to the chemical annotation of the proteins, have larger values for Group 10. Features (27) and (32), with larger values for proteins of Group 10, indicate the

difference between the two Groups structural water atoms allocation and structure. Eventually, features that are connected with the chemical annotations of the interface atoms and the chemical annotations of the surface area, may have greater discriminative power and could consist of classification features. In the second approach, we tried to discriminate the protein Groups, relying to the features we have retrieved and analyzed above. We chose to take into consideration the first Group of each Group family, namely Group 1, Group 3, Group 5, Group 7 and Group 9. Analyzing the normalized mean values, we chose the features that had great value differences in at least one Group. The features are (14),(17),(24),(25),(26),(33),(35),(37),(38) and (40). All these observed features are connected with the chemical annotations of the interface atoms ((14),(17),(24),(25) and (26)) and with the chemical annotations of the surface area ((33),(35),(37),(38) and (40)). Specifically, protein Group (ii) transporter activity proteins, discriminates from the other Groups by the features (14, (17), (31), (33-36) and (38) , having the greatest values in contrast with the other protein families. Protein Group (iv) protein binding transcription factors, discriminates strictly from the other protein families, relying on features (37) and (40) and partly according to the features (24), (25) and (26). Protein Group (v) enzyme regulator proteins, discriminate from the other Groups, relying strictly on the features (24 - 26) and partly on the features (34) and (36). We may conclude in the fact that the chemical annotations of the protein interface and surface provided in the statistic results of the program Intervor, might be classification features between several protein family Groups. Combination of the specific features might also be indicators for classifying a protein, to a protein family.

Bibliography:

Loriot, Sébastien, and Frédéric Cazals. "Modeling macro–molecular interfaces with Intervor." *Bioinformatics* 26.7 (2010): 964-965.

Bouvier, Benjamin, et al. "Shelling the Voronoi interface of protein–protein complexes reveals patterns of residue conservation, dynamics, and composition." *Proteins: structure, function, and bioinformatics* 76.3 (2009): 677-692

Cazals, Frederic. "Revisiting the Voronoi description of protein-protein interfaces: Algorithms." *Pattern Recognition in Bioinformatics*. Springer Berlin Heidelberg, 2010. 419-430.

Cazals, Frederic. “Modeling Macro-Molecular Complexes and Assemblies" *May 2013 presentation @ University of Athens*.

Cazals, Frederic. “Modeling Macro-Molecular Complexes and Assemblies with Voronoi Diagrams" *presentation 2012 @ INRIA Sophia-Antipolis*.

Cazals, Frédéric, et al. "Revisiting the Voronoi description of protein–protein interfaces." *Protein Science* 15.9 (2006): 2082-2092

Janin, Joël, Ranjit P. Bahadur, and Pinak Chakrabarti. "Protein–protein interaction and quaternary structure." *Quarterly reviews of biophysics* 41.02 (2008): 133-180.

Jones, Susan, and Janet M. Thornton. "Principles of protein-protein interactions." *Proceedings of the National Academy of Sciences* 93.1 (1996): 13-20

Bouvier, Benjamin, et al. "Shelling the Voronoi interface of protein–protein complexes reveals patterns of residue conservation, dynamics, and composition." *Proteins: structure, function, and bioinformatics* 76.3 (2009): 677-692.