Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 35 (2014) 1092 – 1101

Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

Graphical models of psychosocial factors in chronic somatic diseases

Anita Galuszka[a], Adam Galuszka[b,*]

[a]*Institute of Psychology, Silesian University Grazynskiego 53, 40-126 Katowice, Poland*
[b]*Institute of Automatic Control, Silesian University of Technology Akademicka 16, 44-100 Gliwice, Poland*

Abstract

In this paper we describe a graph, tree and forest model of psychosocial factors dependencies of chronically ill patients, called graphical models. Foundation of the study was the theory of meaningfulness of suffering by V. E. Frankl. 181 patients with either arterial hypertension or neoplasms with bad prognosis were examined thrice: 0-10 days from the time of diagnosis (stage I), about 5 weeks from the diagnosis (stage II) and at a follow-up about 5 months since stage II (stage III). 75 factors were available for consideration: 17 in stage I, 28 in stage II, 27 in stage III and 4 sets of data that describe populations: age, gender, education, number of stages executed. For both diseases graphs and trees are built under assumption that factors are vertices and significant correlations are edges, leading to model of dependencies between factors. Usefulness of this approach to analysis of difference between diseases is discussed.

Peer-review under responsibility of KES International.

Keywords: Graph methods, neoplasm, hypertension, psychosocial factors

1. Introduction

Two so-called diseases of civilization were taken into account in this paper: arterial hypertension and neoplasm with bad prognosis (Koszarowski 1985, Sheridan and Radmacher 1992). These diseases were explored under 'psychological' point of view, i.e. such psychological properties like: hope, anxiety, meaning of life, coping with stress, different tests of self-esteem and other health cognitive-emotional processes were explored. Most of these factors describe emotional sphere of human life. Such kind of data are usually analyzed using different statistical tests that verify assumed hypothesis (Brzezinski 1997, Ferguson and Takane 1989). In

* Corresponding author.
E-mail address: adam.galuszka@polsl.pl.

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.197

the case explored in this paper data are special since they are unique in country-scale: over 70 psychosocial factors were determined for almost 200 patients suffering from serious diseases. Moreover the dynamic of these disease were taken into account: the data collection has been divided into 3 stages.

A statistical analysis for particular factors from this set of data like hope, anxiety, meaning of life were explored earlier by Galuszka (2005, 2008, 2013).

A motivation for this paper is to consider all of these unique data. Contribution of this paper is a proposition of graph representation for correlation exploration between all available data and forest representation causality relations between factors. As it is shown such model can describe differences and similarities of cognitive-affective processes between diseases and leads to graphical model of dependences between measured factors. Preliminary results were presented in (Galuszka and Galuszka 2009), were graphical representation has been treated as a function of patients gender.

The paper is organized as follow. The data are more precisely explained, next the way of building graphs is described, next the graphs are analyzed. Finally, all is concluded. All figures are made using MATLAB.

2. Data

The data were collected in Upper Silesia hospitals, including the Oncology Center - the M. Skłodowska-Curie Institute in Gliwice, and the Central Clinical Hospital of the Silesian Medical Academy in Katowice. We studied 181 patients: 108 with hypertension and 73 with neoplasm. The dynamics of the disease was taken into account in the research. The patients were examined thrice: stage I, 0-10 days from the time of diagnosis, was regarded as the period of shock due to the patient's learning about his/her severe chronic illness. Negative emotions (mostly anxiety) were expected at that stage, as well as an unfavorable self-rating of subjective health and a low sense of meaning in life. Stage II was carried out about five weeks after the first examination. In that period the patients could be expected to have adapted to their new life situation. They had enough time to undertake health-promoting activities that would improve not only their physical health, but also emotional state. The patients' improved affective condition and involvement in their own treatment should lead to their more favorable self-appraisal of health. Moreover, their perceived meaning in life should be higher than that in stage I. In stage III, about five months from the second examination, the patients could be expected to have markedly adapted to life with their chronic illness. However, various outcomes could be expected – from a significant deterioration of their functioning, through no change to a marked improvement – depending on the course of the disease and progress in its treatment.

Some classical psychological tools were used to transform data to psychosocial factors:

a) Perceived meaning in life was assessed using the Purpose in Life Test (PIL) by J.C. Crumbaugh and L.T. Maholick, in the Polish authorized translation by Z. Płużek. Only part A of the PIL was used, consisting of 20 items with a 7-point rating scale each, on which the respondent is asked to check his/her agreement with the item content (from 7 – fully agree to 1 – disagree). Items concerning death and suicide were excluded as too invasive in patients with a severe physical illness.

b) Two emotional states – hope and anxiety - were measured using The Gottschalk-Gleser Content Analysis Scales (GGCAS). This projective tool allows to measure emotions both at the conscious and unconscious level, taking into account also defense mechanisms. The respondent's current emotional state is assumed to influence his/her perception of the examination situation and his/her choice of the past events. According to the authors, this test yields an interval scale, and so parametric tests can be used in the statistical analysis of the obtained final scores (quoted after Heszen-Niejodek (1987)). In this study inter-rater reliability calculated using Pearson's r was $r = 0.84$ for the Anxiety Scale, and $r = 0.85$ for the Hope Scale.

c) An interview was designed for the purposes of this study to measure the patient's self-appraisal of health. His/her subjective health was rated on a 5-point rating scale, from 5 (very good) to 1 (very bad). The intermediate ratings were: good, middling (average) and bad.

d) Miller Behavior Style Scale (MBSS) by Miller (1987).

e) Medical estimation of health condition.

Finally, the factors are represented by 2 matrices: first collects data for hypertension, second collects data for neoplasm. Both consist of 75 column that correspond to factors. The number of rows corresponds to a number of patients: 108 for hypertension and 73 for neoplasm.

3. Elements of graph theory

In this paragraph we introduce notions of graph theory elements, taken from (Wilson 1998), that are used in the paper.

3.1. Graph

A graph or undirected graph is defined as:

$$G(V,E), \tag{1}$$

where V is a set of vertices, and E is a set of edges that defines which vertices in a graph are connected. A weighted graph is a graph in which each edge is given a numerical weight.

3.2. Tree and maximal spanning tree of graph

Now the tree is introduced. A tree is a set of straight line segments connected at their ends containing no closed loops (cycles). In other words, it is a simple, undirected, connected, acyclic graph. A tree with n vertices has (n–1) graph edges. Conversely, a connected graph with n vertices and (n-1) edges is a tree. Trees with no particular node singled out are sometimes called free trees (or unrooted tree), by way of distinguishing them from rooted trees.

Now a spanning tree is introduced. A spanning tree of a graph with n vertices is a subset of (n-1) edges that form a tree. A maximum spanning tree is a spanning tree of a weighted graph having maximum weight.

3.3. Forest

A forest is an acyclic graph (i.e., a graph without any graph cycles). Forests therefore consist only of (possibly disconnected) trees, hence the name "forest."

3.4. Graph matching

The graph matching problem is very well known and widely explored (e.g. by Eroh and Schultz (1998)). It is known to be computationally difficult i.e. there are not known methods that solve this problem in reasonable time for graphs with relatively big size. In model-based pattern recognition problems, given two graphs the model graph GM and the data graph GD the procedure of comparing them involves to check whether they are similar or not. Generally, the graph matching problem is as follows: given two graphs $GM = (VM, EM)$ and $GD = (VD, ED)$, with the same number of vertices, the problem is to find a one-to-one mapping f:

$$VD \rightarrow VM, \tag{2}$$

such that:

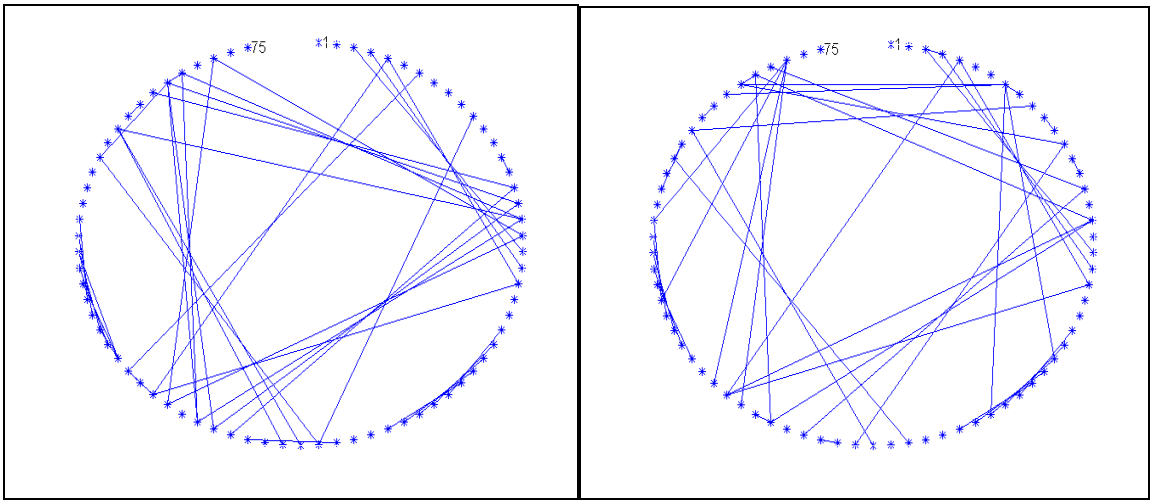
$$(u, v) \in ED \text{ iff } (f(u), f(v)) \in EM \quad (3)$$

When such a mapping f exists, this is called an isomorphism, and GD is said to be isomorphic to GM. This type of problems is said to be exact graph matching.

If two graphs are not isomorphic (exact matching cannot be found) then one can introduce a function that measures similarity between two graphs.

4. Graphs of correlations

It is assumed that a set V is the set that consists of all factors, i.e. graph for each disease consists of 75 vertices (see Appendix A). Nodes from 3 to 19 are for stage I, 20 to 47 for stage II and 48 to 73 for stage III. First and last two describes population properties. It is also assumed that the edge exists when absolute value of Pearson's correlation coefficient between two vertices is greater than 0,5 and significance level is lower than 0,05. The border value of r is chosen arbitrary but is convenient for graph illustration (for lower values of r firstly low correlations are also considered as edges, secondly the number of edges fast increases and the graphs become difficult for presentation).



(a) (b)

Fig. 1. (a) graph for hypertension; (b) graph for neoplasm

The absolute value of Pearson's correlation is needed since the edge denotes only the strength and not the direction of correlation (causalities). Despite the data are of different kind (used scales are: interval and rating) only the correlation is calculated. This is justified since the correlation coefficients will be treated as weights of graphs edges.

The graph for hypertension G1(V1, E1) is shown in the Fig 1a, for neoplasm G2(V2,E2) in the Fig 1b. The vertices of graphs are placed on a circle because our intension is to analyze the dynamic of disease (see section graphs analysis). Summarizing it can be stated that correlation between factors in different stages are modeled by undirected weighted graph.

Graphs were analyzed in the following way:

a) Graph matching problem has been solved to show similarities between two diseases;
b) A hypothetical graph model has been introduced to show differences between two diseases.

4.1. Graphs interpretation

If two graphs are not isomorphic (exact matching cannot be found) then one can introduce a function that measures similarity between two graphs. In the case presented here both graphs are data graphs and the matching problem is reduced to edges matching only since vertices for both disease are the same (are matched):

$$V1 = V2. \tag{4}$$

It leads to proposition of definition of similarity function $F(G1,G2)$ as the number of common edges of graphs $G1$ and $G2$ i.e. size of the intersection over size the sum of sets $E1$ and $E2$:

$$F(G1,G2) = \frac{|E1 \cap E2|}{|E1 \cup E2|}, \tag{5}$$

where $\cap$ and $\cup$ denotes the intersection and the sum of sets, respectively. The minimal value of formula (5) is '0' when there are no common (matched) edges in graphs $G1$ and $G2$ and the maximal value is '1' when graphs $G1$ and $G2$ are identical.

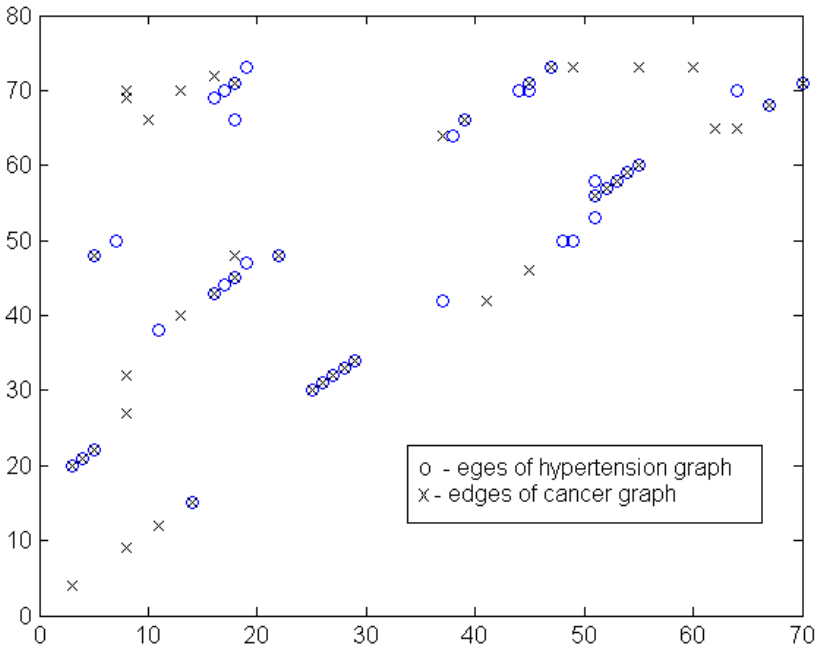


Fig. 2. Illustration for graph matching problem

The edge matching for graphs G1 and G2 is presented in the Fig.2. In the fig.1 axis denotes vertex indices, i.e. point with coordinates X,Y denotes the edge between vertices (factors) X and Y. The number of edges are: 41 for hypertension graph (Fig.1a) and 44 for neoplasm graph (Fig.1b). The number of matched edges (correlations between the same factors for both diseases) is 23 and the size of the sum of edge sets is 60, so:

$$|E1 \cap E2| = 23, |E1 \cup E2| = 60, F(G1, G2) \approx 0.38, \quad (6)$$

So the matching ratio is about 38 %.

4.2. Hypothetical graph

To illustrate differences between diseases a hypothetical model has been introduced in Fig. 3. It is based on assumption that if there are no reasons to disturb factors in time (such reason is a disease), then the same factors measured in different time moments should be strongly correlated. For presented graphs if the same factor in 3 stages of research is correlated then characteristic triangles can be observed in Fig.3. Here as an illustration 3 factors measured in 3 stages has been chosen: a meaning of life (vertices: 5, 22, 48), a hope (vertices: 6, 23, 49) and an anxiety (vertices: 7, 24, 50).

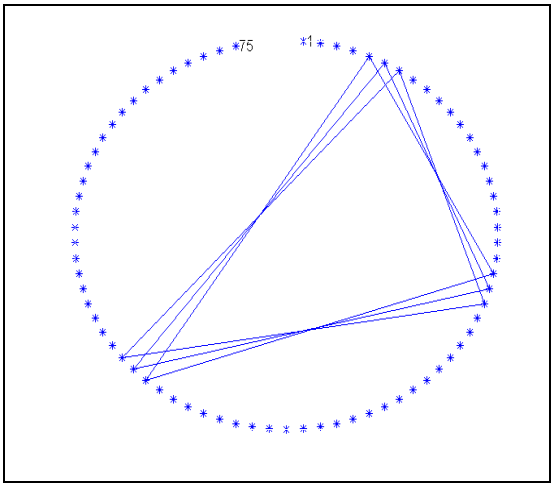


Fig. 3. A hypothetical graph model

A comparison of graphs from figures 1a and 3 results in some similarities (there is a group of triangles for both graphs) whereas a comparison of graph fig. 1b with one from fig. 3 results in many differences: disturbances are observed in triangles regularity. One could introduce scoring function to describe these disturbances precisely.

5. Maximal spanning trees of correlations graphs

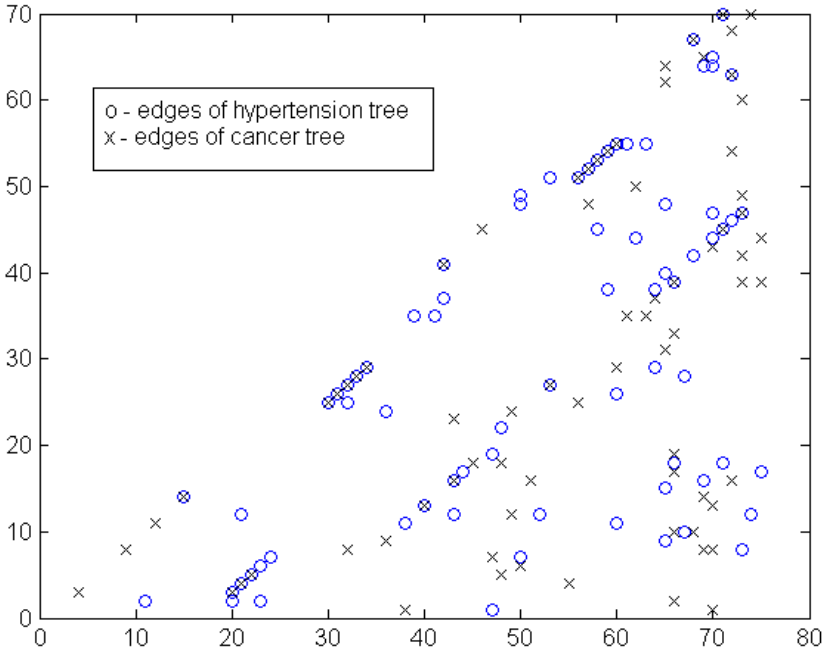


Fig. 4. Illustration of trees matching

Spanning trees construct a sparse sub graph that tells a lot about the original graph. Maximal spanning tree of each graph will illustrate which factors are most connected (correlated). To solve this the minimal spanning tree problem for original graphs with negative weights has been solved. The original graphs has been extended in such a way that set of edged E consists of all significant correlations. This is because as an input 'more' coherent graphs are needed (graphs for $r > 0{,}5$ consists of too many sub-graphs). Maximal spanning tree of hypertension patients data is denoted by $T1$ and for neoplasm patients data by $T2$:

$$T1(TV1, TE1), \quad T2(TV2, TE2), \tag{7}$$

and sets of vertices are the same like for graphs $G1$ and $G2$:

$$V1 = V2 = TV1 = TV2. \tag{8}$$

Edges of maximal spanning trees for both diseases are presented in the Fig.4. The number of edges is now 74 for each tree, and the formula (5) takes the form:

$$F(T1,T2)=\frac{|TE1\cap TE2|}{|TE1\cup TE2|}=\frac{25}{123}\approx 0.2\,. \tag{9}$$

Since correlations does not explain causation relations between factors, in our model there is no distinguished vertex in the tree.

6. Forest of causalities

Directed models are very useful to illustrate cause-effect relations. However, given data one should be very carefully in causality reasoning, e.g. logical implication does not describe causality.

First let us introduce classical definition of cause and effect (Hume 1748): ”We may define a cause to be an object, followed by another, and where all the objects similar to the first, are followed by objects similar to the second.”

Using observations it is convenient to use probabilistic approach for causality reasoning. Probabilistic theories of causation is that causes raise the probability of their effects; an effect may still occur in the absence of a cause or fail to occur in its presence. Thus smoking is a cause of lung cancer, not because all smokers develop lung cancer, but because smokers are more likely to develop lung cancer than non-smokers (Eells 1991).

Very important thing is that correlation does not imply causation. The statement: if A occurs in correlation with B. implies that A causes B is a logical fallacy because there are at least four other possibilities:

- B may be the cause of A, or
- some unknown third factor is actually the cause of the relationship between A and B, or
- the ”relationship” is so complex it can be labelled coincidental (i.e., two events occurring at the same time that have no simple relationship to each other besides the fact that they are occurring at the same time).
- B may be the cause of A at the same time as A is the cause of B (contradicting that the only relationship between A and B is that A causes B). This describes a self-reinforcing system, i.e. systems with positive feedback.

In case of graphs *G1* and *G2* the only directed correlations representing cause-effect relations are correlations between vertices (factors) measured in different stages. Basing on this assumption it is possible to construct directed forests *F1* for hypertension data and *F2* for neoplasm data, presented in Fig.5.

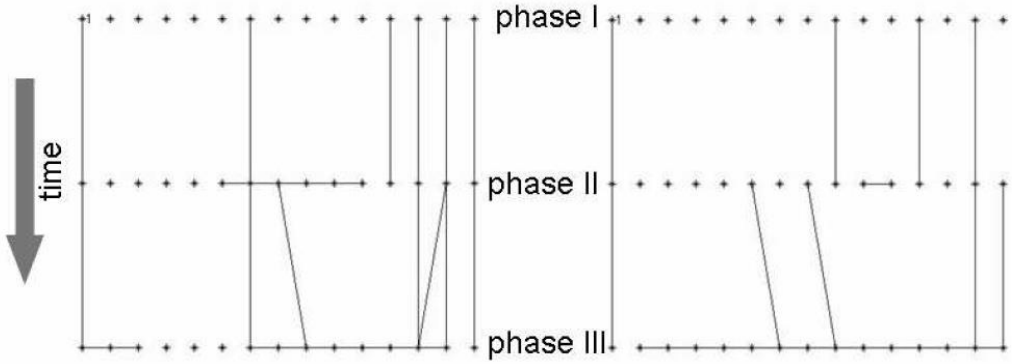


Fig. 5. Forests *F1* and *F2* of causality relations

7. Conclusion

In this paper a graph based approach is proposed to analysis a set of psycho-social factors. The graph analysis results in some conclusion. Firstly should be noted that diseases of different kind differently influences on cognitive, emotional and behavioral sphere of a man (differences in corresponding to this spanning trees – only 20 % of matching) but there are also many similarities (factors correlations are in 38 % matched). The proposed hypothetical model illustrate how dramatic is the influence of serious disease (neoplasm) on psychosocial factors comparing to the second researched disease (hypertension): the regularities in graphs are much more disturbed in the first case (see 1a vs 1b).

The issue of psycho-social factors of patients with a chronic illness is of particular importance. Suffering associated with illness is destructive in itself. Finding meaning in suffering, and thus the meaning in his/her whole existence, helps the patient to adapt to new conditions, to gain a better quality of life, and to orient himself/herself towards supra-values.

Since all graphical models described here are built using real patients data, they can be applied as feedback information that indicates which medical and psychological interferences should be changed in order to improve patients quality of life. Similarity functions for graphs and trees proposed in the paper together with scoring function of graph regularity can be apply to description of differences between psycho-social state of patient groups suffering from different illnesses.

In the future research more complex graph analysis methods to describe these psycho-social factors are planned to be applied. We hope that graph models can stay an useful tool for analysis of coping with disease.

Acknowledgements

The research is financed by Institute of Psychology BK funds for the first author and Institute of Automatic Control BK grant No. 214/RAu1/2013 for the second author in the year 2014. Collection the data has been supported by KBN grant 1 P 106 004 05 under supervision of Prof. I. Heszen-Niejodek.

References

[1] Brzezinski J. (1997). *Methodology of psychological research*, Warsaw, PWN – Polish Scientific Publishers, 1997.
[2] Eells, E. (1991). *Probabilistic Causality*. Cambridge, U.K.: Cambridge Univ. Press.
[3] Eroh L. and M. Schultz. (1998). Matching graphs. *Journal of Graph Theory*, 29(2):73-86.
[4] Ferguson G.A., Y.Takane (1989). *Statistical Analysis in Psychology and Education*, New York, McGraw-Hill, Inc., 1989.
[5] Galuszka Anita. (2006): Coping with chronic somatic diseases - arterial hypertension and neoplasm, *Psychology & Health* Volume: 21: 53-53.
[6] Galuszka, A. (2008). Cognitive-affective correlates of meaningfulness in life in chronic somatic disease. *Computerized Content Analysis of Speech and Verbal Text and its Many Applications*, 43-55.
[7] Galuszka, A. (2013). Personality and social resources in case of serious somatic disease. *Psychology & Health,* Vol 28: 217-217.
[8] Galuszka, Adam., Galuszka, Anita. (2009). Graph analysis of psychosocial factors correlations of patients with hypertension and neoplasm in Upper Silesia, Poland, *Proc. 9th WSEAS Int. Conf. Simulation, Modelling and Optimization, SMO '09, 5th WSEAS Int. Symp. Grid Computing, Proc. 5th WSEAS Int. Symp. Digital Libraries, Proc. 5th WSEAS Int. Symp. Data Mining PP*. 375 -380.
[9] Heszen-Niejodek I. (1987). Skale Analizy Treściowej Gottschalka i Gleser, [Content Analysis Scales by Gottschalk and Gleser], *Psychological Review* (in polish), 30, 1987, 987-1002.
[10] Hume D. (1748). An Enquiry Concerning Human Understanding, *Harvard Classics* Volume 37, Copyright 1910 P.F. Collier and Son.
[11] Koszarowski T. (1985). *Onkologia kliniczna*, [Clinical Oncology], Warsaw, PZWL, 1985.
[12] Papadimitriou Ch. (2001). *Teoria złożoności obliczeniowej*, [Theory of the Complexity], PWN, Warsaw.
[13] Sheridan C.L., S.Radmacher (1992). *Health Psychology. Challenging the Biomedical Model*, New York, John Wiley & Sons.
[14] Wilson R.J. (1998). *Wprowadzenie do teorii grafów*, [Introduction to the Graph Theory] , PWN, Warsaw.

[15] Miller S.M. (1987). Monitoring and Blunting. Validation of Questionnaire to Asses Styles of Information Seekeing under Threat, *Journal of Personality and Social Psychology*, 52, 345-353.

Appendix A. List of psychosocial factors that are graphs vertices.

vertex	1st stage factors	vertex	2nd stage factors	vertex	3rd stage factors
1	age	20	monitoring	48	PIL- Purpose in Life Test (meaning in life)
2	education	21	blunting	49	fear
3	monitoring	22	PIL- Purpose in Life Test (meaning in life)	50	hope
4	blunting	23	fear	51	home/family duties
5	PIL- Purpose in Life Test (meaning in life)	24	hope	52	satisfaction of family duties
6	fear	25	home/family duties	53	appraisal of family relations
7	hope	26	satisfaction of family duties	54	professional duties
8	home/family duties	27	appraisal of family relations	55	satisfaction of professional duties
9	satisfaction of family duties	28	professional duties	56	plans for future
10	appraisal of family relations	29	satisfaction of professional duties	57	social activity
11	professional duties	30	plans for future	58	appraisal of social activity
12	satisfaction of professional duties	31	social activity	59	global activity
13	plans for future	32	appraisal of social activity	60	symptoms of illness
14	social activity	33	global activity	61	self appraisal of health
15	appraisal of social activity	34	symptoms of illness	62	medical appraisal of health
16	global activity	35	self appraisal of health	63	Kat I- searching for information
17	symptoms of illness	36	medical appraisal of health	64	Kat II - adherence to medical recommendations
18	self appraisal of health	37	Kat I- searching for information	65	Kat III – health activity
19	medical appraisal of health	38	Kat II - adherence to medical recommendations	66	Kat IV - avoiding harmful activities
		39	Kat III – health activity	67	Kat V- defense mechanisms
		40	Kat IV - avoiding harmful activities	68	S I - self searching for information
		41	Kat V- defense mechanisms	69	S II - self adherence to medical recommendations
		42	S I - self searching for information	70	S III - self health activity
		43	S II - self adherence to medical recommendations	71	S IV - self avoiding harmful activities
		44	S III - self health activity	72	S V - self defense mechanisms
		45	S IV - self avoiding harmful activities	73	subjective appraisal of changes of health
		46	S V - self defense mechanisms	74	patient number
		47	subjective appraisal of changes of health	75	sex