A Graph Mining Perspective on Graphlet-Based Network Similarity

MARTIN NEHÉZ
Slovak University of Technology in Bratislava
Bratislava, Slovak Republic

MAREK LELOVSKÝ
Slovak University of Technology in Bratislava
Bratislava, Slovak Republic

Abstract: Analysis of networks has emerged in recent years as an important knowledge management tool. Specifically, the social network analysis can positively influence customer behavior prediction, identification of proper groups of shoppers/clients or efficiency of viral advertising. In particular, mining of graphlets (i.e. small induced subgraphs) has become a prominent research topic. Although it originally comes from bioinformatics, it finds considerable applications in social networks, as well. In this paper, we focus on the network similarity problem and related graphlet-based algorithms and corresponding data management processes, respectively. We describe known solutions, draft their possible alternatives and pose several open problems.

Keywords: networks; protein-protein interactions; graphlets; similarity measures.

1 Introduction

The basis of the network science was grounded by A. L. Barabási and his research team approximately 20 years ago. It is a field which primarily deals with a structure and dynamics of real-world networks and its research methods are taken over from statistical physics, combinatorics, computer science and probability theory. The most remarkable applications of the network science were found in electrical engineering, informatics, operations research (e.g. transportation), social science, bioinformatics, security and cyberwarfare. Overlapping of the network science and knowledge management is growing in significance more and more. As an example, one can mention results regarding customer behavior predictability in social networks [13], community detection [14, 19], studying the dynamics of viral marketing [11], etc. Another challenging subject is the graphlets mining problem, which originated in bioinformatics [16]. Therefore, there is a close relationship among the network science, bioinformatics and the knowledge management. Details are discussed in the monograph by I. Jurišica and D. Wigle [10]. A typical example represents an analysis of protein-protein interactions.

In organisms on a subcellular level, major biological processes are provided by biochemical or functional interactions among macromolecules, such as proteins. Specifically, many of the key biological activities (e.g. metabolism, gene expression, immunity, signaling) are mediated through protein interactions [18]. For instance, there are nearly half a million protein interactions in a human body but so far, only a part of them has been investigated in detail [18]. In order to extract valuable knowledge concerning these processes, a network-based abstraction is employed. Corresponding networks are called protein-protein interaction networks, shortly PPI networks. Studying them is currently one of the prominent topics in bioinformatics.

Examination of similarities among PPI networks of the same type or, on the contrary, finding anomalies among them is an appealing task, which has found applications in biomedicine. A

famous algorithmic method for this purpose was invented by N. Pržulj [16, 17]. It is based on frequency analysis of small patterns (called graphlets) occurring in networks. During 15 years of practical usage of Pržulj's method, it proved its efficiency, but several weaknesses were observed, as well. In this paper, we address two early emerged limitations, which are probably the most significant ones. Although there was a lot of effort dedicated to eliminating them, the discussion on this theme is still going on.

- The first limitation comes from the fact that for real-world PPI networks with thousands of nodes, algorithmic counting of graphlet frequencies requires high demand on computational resources. Fortunately, due to recent highly nontrivial graph-theoretical results (see e.g. [7]), graphlets counting algorithms were improved considerably. Known software programs are e.g. FANMOD, GraphCrunch, RAGE (surveyed in [7]). Currently, the most powerful one seems to be ORCA – the Orbit Counting Algorithm, which was developed by T. Hočevar [7]. Although the ORCA performs very well using even low-cost hardware, research and development of other new methods and software programmes is permanently in progress [3].
- The second limitation regards the statistical methods measuring networks similarity. The original idea considered visual comparison of two graphlet frequency distributions (each for a different network) but was enriched by the usage of so called *relative graphlet frequency distance* [16]. Unfortunately, it seems that the relative graphlet frequency distance is very context-sensitive. Moreover, T. Hočevar said [8]: "*So far I haven't seen a fixed threshold used for deciding whether two networks are similar or not*". A new systematic measure was defined in [17], which is *graphlet degree distribution agreement*, shortly GDD agreement. Such a measure is used frequently and seems to be better suited for measuring the networks similarity than the former one. Nevertheless, there exists space for further research in this area.

In this paper, we turn our attention towards the network similarity comparison process using the ORCA software. Moreover, in Section 3, we introduce two quantities which are originated in statistical divergence theory. They are the total variation distance and the Hellinger distance, respectively. These quantities are used as network similarity measures in our experimental study (Section 4). In order to conduct experimental simulations, we describe the ORCA-based workflow, which is utilized in the dataset processing. In Section 4, we argue why the usage of the new measures is reasonable. We discuss their advantages and weaknesses and suggest the directions for future research.

2 PPI Networks and Graphlets in Bioinfomatics

2.1 Graph Theory

Roughly speaking, a network is a collection of nodes with connections among them. Formally, a *graph* (i.e. an equivalent notion to the network) is a pair G=(V, E), where the set V represents a collection of *vertices* and the set E comprises *edges*. In reality, vertices represent nodes of a given network and links (or connetions) are modelled by edges. We assume that each edge connects exactly two vertices as its endpoints. Moreover, links (edges) are without directions and they do not form "self-loops". Between ech pair of vetices are no multiple edges. These assumptions match all types of networks which we are dealing with in this paper. Given a graph G=(V, E), a *subgraph* of G is a graph H=(W, F) such that $W \subseteq V$ and $F \subseteq E$, i.e. both W

and F are subsets of V and E, respectively. Induced subgraphs are instances of subgraphs which are often used in bioinformatics. Thier definition is as follows. Let G=(V, E) be a graph and let $W \subseteq V$. An *induced subgraph* of G with respect to W is the graph whose vertex set is W and whose edge set consits of all such edges of E that have both endpoints in W. A *connected graph* is a graph in which all pairs of distinct vertices are connected by a path, i.e. each vertex is reachable from another one. An equivalence relation between two graphs is defined via isomorphis. Two graphs $G_1=(V_1, E_1)$, $G_2=(V_2, E_2)$ are said to be *equivalent* if there is a one-to-one mapping $\sigma: V_1 \rightarrow V_2$, called *isomorphism*, such that for all pairs $u, v \in V_1$ it holds $\{u, v\} \in E_1$ if and only if $\{\sigma(u), \sigma(v)\} \in E_2$. An isomorphism of a graph G onto itself (i.e. $\sigma: G \rightarrow G$) is said to be *automorphism*. Given a graph G, let Aut(G) denote a group of all automorphisms $\sigma: G \rightarrow G$. Let $v \in V(G)$ be a vertex, an *orbit* of v is a set of all images $u=\sigma(v)$ for all automorphisms $\sigma \in Aut(G)$. Examples of orbits are described at the end of this subsection.

Let G = (V,E) be a graph, a *graphlet* is a connected induced subgraph of G with at most *5* vetices. Two graphlets are the same if there exists an isomorphism such that it maps one graphlet to the other one. (Two different occurrences of the same graphlet are usually referred to as its copies.) Graphlets play a seminal role in bioinformatics [10, 16, 17]. There are totally *30* graphlets with *2, 3, 4* and *5* vertices. Ordering and labeling of graphlets with at most *4* vetices is shown in **Fig. 1**. All graphlets are denoted by $g_1, g_2, ..., g_{30}$. Note that e.g. g_7 has three different orbits (two vertices with degree *2* belong to the same orbit, the central vertex forms another single-element orbit and the single vertex with degree *1* represents the third orbit) and all vertices of g_3 belong to the same *3*-element orbit (similarly, g_6 has the only one *4*-element orbit). Moreover, g_2, g_4, g_5 and g_8 have equally two orbits.

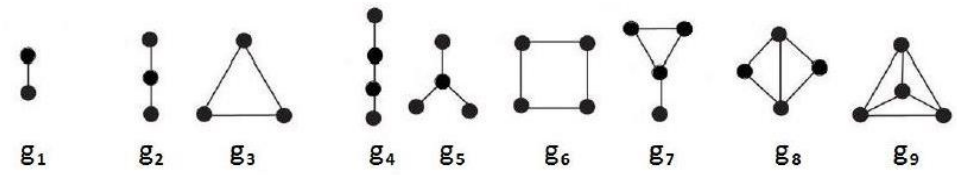

Fig. 1 *All graphlets with at most 4 vertices.*

2.2 Proteomics and Networks

Proteins are macromolecules of aminoacids which are essential components in all organisms. They do not act in isolation but they interact among each other [6]. Protein-protein interactions (PPI) are neither static nor stable;; instead, they are dynamic. Some of them are quick but others are slow. Clearly, in PPI networks, vertices represent proteins and edges interactions. From a general point of view, PPI networks represent a type of biological networks which are commonly called interactome networks [9]. Their structure and behavior is very complex, e.g. PPI networks of mammals have approximately ten thousand proteins and hundreds of thousands interactions. An example of the PPI network Caenorhabditis Elegans is shown in **Fig. 2**. It consists of *2903* proteins and *4631* interactions.

Detection of protein-protein interactions is also highly nontrivial and it encompasses various methods (indirect, in vivo, in vitro, etc.) [6]. PPI networks share some common properties. Specifically, empirical evidence that the structure of PPI networks is close to geometric random graphs was published in [16]. To obtain this result, the graphlet-based similarity method was

adopted. This and further knowledge concerning structure of PPI networks provides a basis which may be essentially helpful in drug design or in related biomedical applications [1, 5, 6].

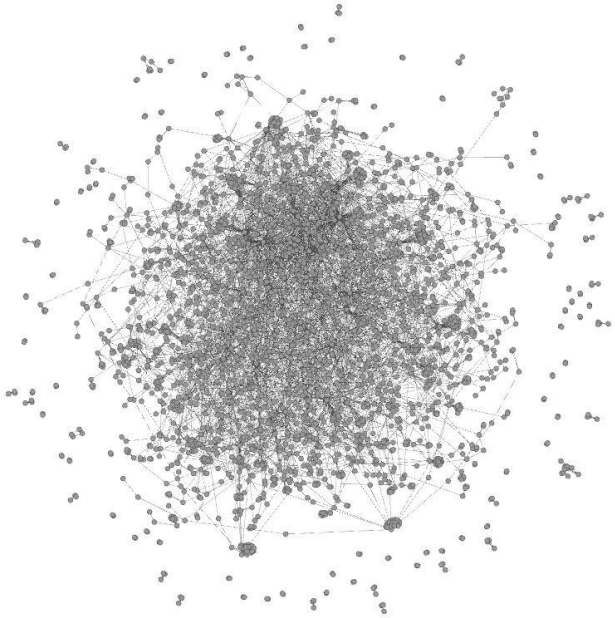

Fig. 2 *PPI network of the roundworm Caenorhabditis Elegans (2903 proteins and 4631 interactions) drawn by the freeware Gephi [2]. The dataset was downloaded from [15].*

3 Networks Similarity Measures

There is a wide variety of approaches to comparing networks or graphs. These approaches range from exact and very strict (graph isomorphism) through various kinds of equivalences (automorphic, regular, etc.) to statistical comparison methods. The latter ones are based on the so-called structural similarity, i.e. an approach in which graphlets occurrence[1] in both compared networks is evaluated by various statistical measures [16, 17]. A selection of these measures (or quantities) is listed in the following subsections.

3.1 Relative Graphlet Frequency Distance

Let G=(V, E) be a graph. For i=1,…, 30, let $N_i(G)$ denote the number of graphlets g_i in a graph G. Let T(G) denote the total number of graphlets in G, i.e.

$$T(G) = \sum_{i=1}^{30} N_i(G) \tag{1}$$

and let the *negative logarithmic relative frequency* of a graphlet g_i be as follows

[1] Or alternatively, subgraphs occurrence

$$F_i(G) = -\log\left(\frac{N_i(G)}{T(G)}\right). \quad (2)$$

Given two graphs G and H, the *relative graphlet frequency distance* (or *distance for brevity*) is defined as follows.

$$D(G,H) = \sum_{i=1}^{30} |F_i(G) - F_i(H)| . \quad (3)$$

Recall that the graphlet g_1 (i.e. a single edge) is a trivial case. Therefore, it occasionally suffices to set i=2 in all lower bounds of summations in the above formulas.

Two major limitations of the relative graphlet frequency distance are as follows.

1. If $N_i(G)$ is zero then the corresponding value $F_i(G)$ is undefined in equation (2). It causes that all undefined values of both $F_i(G)$ and $F_i(H)$ had to be omitted in equation (3). Such a fact may influence negatively the accuracy of the relative graphlet frequency distance *D(G, H)*.
2. It is difficult to determine bounds on *D(G, H)* generally. It follows that for a computed value of *D(G, H)*, it is questionable whether graphs in the question are "very similar", "slightly similar" or "not similar". In other words, there is no widely accepted threshold on relative graphlet frequency distance which could be used for deciding whether graphs G and H are similar or not.

The relative graphlet frequency distance was introduced in [16]; however, due to the above reasons, other similarity measures were defined in later works of N. Pržulj at al. They are e.g. graphlet degree distribution agreement (shortly GDD agreement), graphlet degree vectors and signature similarity (see [17] for the details). Due to the space limitations, these quantities are not mentioned in this paper. Instead, we suggest two quantities which are described bellow. Their origin comes from the theory of statistical divergence.

3.2 Total Variation Distance

According to the previous results, the problem of similarity between two networks can be formulated, in essence, as a problem of comparison for two statistical distributions. Such a problem is usually solvable by the statistical methods, namely by statistical divergence. Only the discrete case is useful for our purpose. For two discrete distributions $P = (p_i)_{i=1}^n$ and $Q = (q_i)_{i=1}^n$, the statistical divergence is a "cumulative" quantity which is proportional to the sum of distances between all pairs (p_i, q_i) for i=1, ..., n. In this paper, we use two discrete cases of such a quantity: the total variation distance and Hellinger distance. In order to measure the graph similarity, both are modified accordingly.

Let G, H be two graphs. Recall that $N_i(G)$ and $N_i(H)$ are the numbers of graphlets g_i in a graph G (or in H, respectively) for i=1, …., 30. Recall also that T(G), i.e. the total number of graphlets in G, is expressed by eq. (1) and the same expression holds for T(H) as well. The *total variation distance of graphs* G and H is defined as follows

$$\delta(G,H) = \frac{1}{2}\sum_{i=1}^{30}\left|\frac{N_i(G)}{T(G)} - \frac{N_i(H)}{T(H)}\right|. \quad (4)$$

3.3 Hellinger Distance

Although it is based on the same idea, such a measure is more complex than the previous one. All symbols have the same meaning as above. The *Hellinger distance of graphs* G and H is defined as follows:

$$HD(G,H) = \left[\frac{1}{2}\sum_{i=1}^{30}\left(\sqrt{N_i(G)/T(G)} - \sqrt{N_i(H)/T(H)}\right)^2\right]^{1/2}. \quad (1)$$

It is easy to see that if two graphs are similar, then the value of HD is close to zero (and nonnegative); otherwise it is approaching to one. Therefore, such a measure provides a sufficient tool whenever one needs to distinguish "degree of similarity" between two graphs. Moreover, a deeper insight can confirm that both major limitations of the relative graphlet frequency distance are eliminated in the Hellinger distance.

Within the rest of this paper, the relative graphlet frequency distance, the total variation distance and Hellinger distance are commonly called as *distances*.

4 A Case Study

In this section, we demonstrate the graphlet-based network comparison method using the ORCA software tool [7]. However, the decision regarding the ORCA utilization caused some specific problems which had to be solved. Namely, due to the fact that ORCA requires a strict format of input data (and additionally, the output is a large matrix of integers), it was necessary to design a new workflow (i.e. the knowledge dicovery process) suitable for our purpose. The details are explained below.

4.1 Workflow and Datasets under Study

Our workflow is derived from the generally accepted knowledge discovery process described in [10], pp. 3-6. It represents the transformation of data to knowledge and it is suitable for various domains, including biological knowledge discovery applications, as well. It usually involves 5 steps: selection, preprocessing, transformation, data mining and interpretation. Our workflow is its modification and consists of the following steps.

1. Data preprocessing. The input format of files processed by ORCA should satisfy two necessary requirements: there are no multiple edge occurrences, and labels of all vertices are ordered consecutively, using values from 0 to n-1. To do so, we use two short programs (in Matlab and Python, respectively) which were designed in [12]. These programs are able to do the transformation of input data as it is desirable.
2. The graphlet frequency counting using ORCA. This part represents the first step of the knowledge discovery process. The ORCA was downloaded from URL: http://www.biolab.si/supp/orca/, section "*Download*"; the user's manual can be found on the same web page and the implementation details were published in [7]. For a given graph with n vertices, the output from ORCA is the file which contains

a matrix of n×73 integers. Its value $a_{i,j}$ is the frequency of *i*th vertex in the *j*th orbit, where i=0, …, n-1 and j=0, …, 72 (see [7] for details).

3. Statistical analysis. We used the MS Excel for this purpose. The output file was imported to MS Excel (in "csv"-format) and analyzed according to the measures described in Sect. 3.
4. Interpretation. Four different criteria of similarity comparison were used (see Sect. 3). Such an approach enables to compare their abilities (or weaknesses) and provides a basis for new findings, as well.

The above steps were used to evaluate the similarity of three sample networks which were generated artificially. Datasets were downloaded from http://www.biolab.si/supp/orca/, section "*Download*". They represent three different networks, whose elementary properties are listed in **Tab. 1**. The last column of the table (The Model) refers to the graph model which was used to generate the corresponding network. Visual representation of the networks was drawn by the freeware Gephi [2], see **Fig. 3**.

Tab. 1 *Properties of 3 sample networks.*

Denotation	Number of vertices	Number of edges	The Model
ba_1k_2k.in	1000	1996	Barabási-Albert
ba_1k_4k.in	1000	3984	Barabási-Albert
geo_1k_4k.in	1000	4000	Geometric random graphs

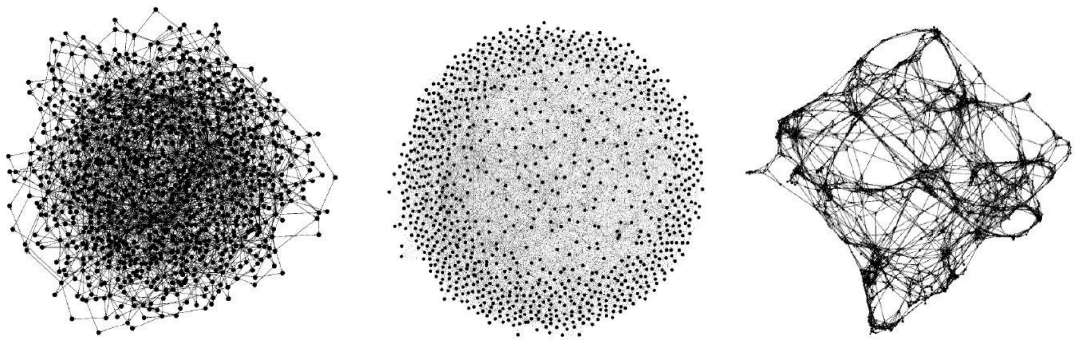

Fig. 3 *The networks ba_1k_2k, ba_1k_4k and geo_1k_4k (from the left to right) drawn by the freeware Gephi [2].*

4.2 Results and Intrepretation

The similarity of above networks was evaluated by four different ways. The first one is based on a visual comparison of graphlet frequencies, three others are based on similarity measures introduced in Sec. 3, i.e. distances.

Recall that each output generated by ORCA comprises of a n×73 matrix ($a_{i,j}$) with vertex-orbit frequencies. In order to compute $N_i(G)$, i.e. the number of graphlets g_i (for i=1,…, 30) in a given graph G, we summed elements of the matrix ($a_{i,j}$) in its *j*th column iff the *j*th orbit corresponds to the graphlet g_i. Such a sum, divided by the number of occurrences of *j*th orbit in g_i, equals to $N_i(G)$. For each nonzero value $N_i(G)$, the negative logarithmic relative frequency

of a graphlet g_i (i.e. $F_i(G)$) was computed according to equation (2). If $N_i(G)=0$ then $F_i(G)$ is undefined. The frequency graphs of values $F_i(G)$ for all three networks are shown in **Fig. 4**. The values in all cases when $N_i(G)=0$, are omitted. One can see that frequencies of the pair *ba_1k_2k* and *ba_1k_4k* are more resembling than of other pairs.

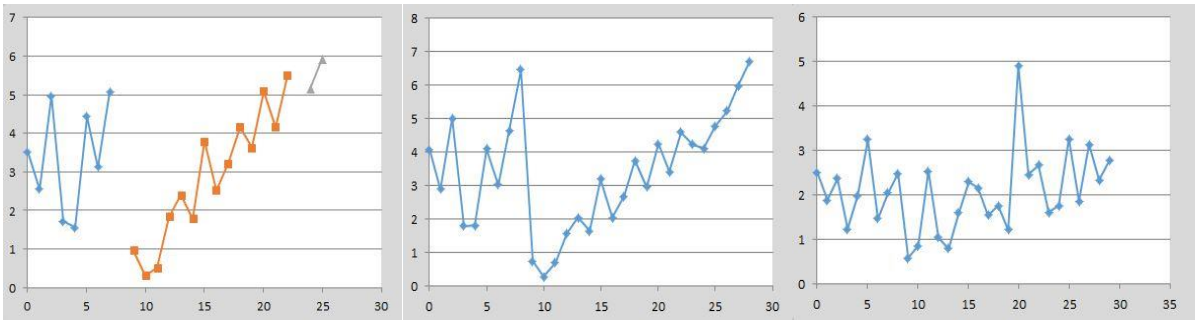

Fig. 4 *Frequencies of $F_i(G)$ for three networks (ba_1k_2k on the left, ba_1k_4k in the middle and geo_1k_4k on the right). Numbers of graphlets (1,..., 30) are in x-axis, values of $F_i(G)$ are in y-axis.*

According to equation (3), the relative graphlet frequency distance *D* was computed for each pair of networks. In corresponding summations, all indices i, for which $|F_i(G)-F_i(H)|$ are undefined, are omitted. The computed values of the relative graphlet frequency distance are listed in the second column of **Tab. 2**. The numbers of all defined values which contributed to the resulting sum (see eq. (3)) are in parentheses in the same column. Their numbers are: 24, 24 and 29, respectively. (Undefined values of $|F_i(G)-F_i(H)|$ are not included.) Values of the total variation distance δ and the Hellinger distance *DH* were computed by equations (4) and (5), respectively. The number of contributed summands were always 30 in both of these distances. The resulting values are listed in the 3rd and the 4th column of **Tab. 2**, respectively.

Tab. 2 *Values of three similarity measures for all pairs of compared networks. Numbers of summands are in parentheses for the relative graphlet frequency distance. (As regards the other two measures, the numbers of summands are always 30.)*

Compared networks	**Relative graphlet freq. distance *D***	**Total variation distance δ**	**Hellinger distance *HD***
ba_1k_2k *vs.* ba_1k_4k	10.67 (24)	0.1265	0.1211
ba_1k_2k *vs.* geo_1k_4k	35.69 (24)	0.6777	0.6166
ba_1k_4k *vs.* geo_1k_4k	45.87 (29)	0.5814	0.5444

The visual comparison of networks's frequency distributions (**Fig. 4**) is only an auxiliary criterion. More significant knowledge can be obtained by the usage of distances (see **Tab. 2**). In general, if a distance is smaller, then the similarity of compared networks is more expressive (and vice versa). As it is shown in **Tab. 2**, the networks *ba_1k_2k*, *ba_1k_4k* are the most similar out of all pairs. On the other hand, the distances of pairs in which the *geo_1k_4k* is occurred are essentially greater. Note that values of both distances δ and *HD* are the greatest for the pair (*ba_1k_2k*, *geo_1k_4k*), which does not correspond to the value of the distance *D*.

Therefore, the accurate judgment can not be formulated. It is only possible to say that the similarity of all pairs of networks in which *geo_1k_4k* occurs is not significant.

4.3 Discussion

Despite the small number of samples in our dataset, the obtained results are beneficial. Both major weaknesses of the relative graphlet frequency distance have been confirmed in our study. These weaknesses, however, do not occur in the total variation distance and the Hellinger distance.

The similarity of the above networks was evaluated in four different ways. Although they always lead to a similar conclusion, we suggest to prefer the total variation distance and the Hellinger distance, respectively. The total variation distance is, in a sense, a modification of the relative graphlet frequency distance but the Hellinger distance represents a more sophisticated quantity. The authors believe that due to its suitable properties, the Hellinger distance could represent a reference similarity measure. However, such an argument needs to be verified in further experimental work.

5 Conclusions

The paper is focused on selected aspects of the graphlet-based similarity analysis of networks. We address the problem of an appropriate networks similarity measure, which has been already discussed in [17]. Our contribution is based on computer simulations. In order to conduct them, the ORCA-based workflow was described and used. Comparing three network similarity measures leads to the observation that we recommend the Hellinger distance as the most suitable one.

In order to verify our findings, it would be desirable to enlarge samples dataset. However, additional experimental studies require an improvement of the ORCA-based workflow. Due to this reason, the authors are currently working on a new software which would perform the processing of statistical analysis more efficiently. Another future research could be aimed at new network similarity measures design.

Acknowledgment

The authors are grateful to prof. M. Terek for his valuable comments and suggestions on the conference presentation. The work of M. Nehéz is supported by the Slovak Science Grant Agency under the grant VEGA 1/0026/16.

References

1. ATHANASIOS, A., CHARALAMPOS, V., VASILEIOS, T. and ASHRAF, G.M., 2017. Protein-Protein Interaction (PPI) Network: Recent Advances in Drug Discovery. *Current Drug Metabolism*, 18(1), pp. 5-10.
2. BASTIAN, M., HEYMANN, S. and JACOMY, M., 2009. Gephi: an open source software for exploring and manipulating networks. In: ICWSM 2009, *3rd Int. AAAI Conference on Weblogs and Social media*. 2009. The AAAI Press, pp. 361-362.
3. CHEN, X. and LUI, J.C.S., 2018. Mining Graphlet Counts in Online Social Networks. *ACM Transactions on Knowledge Discovery from Data*, 12(4), pp. 41:1-41:38.
4. DANG, J., HEDAYATI, A., HAMPEL, K. and TOKLU, C., 2008. An ontological knowledge framework for adaptive medical workflow. *Journal of Biomedical Informatics*, 41(5), pp. 829-836.
5. FENG, Y., WANG, Q. and WANG, T., 2017. Drug Target Protein-Protein Interaction Networks: A Systematic Perspective. *BioMed Research International*, 2017, Article ID 1289259, pp. 1-13.
6. GONZALEZ, M.W. and KANN, M.G., 2012. Chapter 4: Protein Interactions and Disease. *PLOS Computational Biology*, 8(12), e1002819, pp. 1-11.
7. HOČEVAR, T. and DEMŠAR, J., 2014. A combinatorial approach to graphlet counting. *Bioinformatics*, 30(4), pp. 559-565.
8. HOČEVAR, T., 2018. *Personal Communication*. [May 10, 2018].
9. JORDÁN, F., NGUYEN, T.P. and LIU, W.C., 2012. Studying protein-protein interaction networks: a systems view on disease. *Briefings in Functional Genomics*, 11(6), pp. 497-504.
10. JURISICA, I. and WIGLE, D., 2006. *Knowledge Discovery in Proteomics*. Boca Raton: CRC Press, Taylor & Francis Group.
11. LESKOVEC, J., ADAMIC, L.A. and HUBERMAN, B.A., 2007. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1), Article No. 5.
12. KARAFFOVÁ, K., CHOWANIECOVÁ, D., GUBRICKÝ, M. and MALOVEC, T., 2018. Methods for Graphlets Enumeration. Semestral project (in Slovak), Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava.
13. OSTROWSKI, D.A., 2012. *Social network analysis for consumer behavior prediction*. Available at: <http://worldcomp-proceedings.com/proc/p2012/ICA3445.pdf> [Accessed August, 10 2018].
14. PAN, G., ZHANG, W., WU, Z. and LI. S., 2014. Online Community Detection for Large Complex Networks. *PLOS ONE*, 9(7), e102799.
15. Protein-protein interaction (PPI) networks and Gene Ontology annotation files. *Datasets*. [online] Available at: <https://www.comp.nus.edu.sg/~whsu/IRAP/datasets.html> [Accessed September 5, 2018].
16. PRŽULJ, N., CORNEIL, D.G. and JURISICA, I., 2004. Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18), pp. 3508-3515.
17. PRŽULJ, N., 2007. Biological Network Comparison Using Graphlet Degree Distribution. *Bioinformatics*, 23(2), pp. 177-183.
18. SEVIMOGLU, T. and ARGA, K.Y., 2014. The role of protein interaction networks in systems biomedicine. *Computational and Structural Biotechnology Journal*, 11(18), pp. 22-27.

19. TANG, L. and LIU, H., 2010. *Community Detection and Mining in Social Media*. 1st ed. Williston: Morgan and Claypool Publishers.

Contact data:

RNDr. Martin Nehéz, PhD.
Institute of Information Engineering, Automation and Mathematics
Faculty of Chemical and Food Technology
Slovak University of Technology in Bratislava
Radlinského 9, 812 37 Bratislava, Slovak Republic
martin.nehez@stuba.sk or nehez841@gmail.com

Ing. Marek Lelovský
Department of Languages
Faculty of Chemical and Food Technology
Slovak University of Technology in Bratislava
Radlinského 9, 812 37 Bratislava, Slovak Republic
marek.lelovsky@stuba.sk