Optimizing Social Interaction
A Computational Approach to Support Patient Engagement

Italo Zoppis[1], Riccardo Dondi[2], Eugenio Santoro[4],
Gianluca Castelnuovo[3], Francesco Sicurello[1] and Giancarlo Mauri[1]
[1]*Department of Computer Science, University of "Milano-Bicocca", Milano, Italy*
[2]*Department of Letters, Philosophy, Communication, University degli Studi di Bergamo, Bergamo, Italy*
[3]*Department of Psychology, University "Cattolica del Sacro Cuore", Milano, Italy*
[4]*Laboratory of Medical Informatics, Department of Epidemiology, IRCCS, Mario Negri, Milano, Italy*

Keywords: Social Networks, Optimization, Cohesive Sub-Graphs, Genetic Algorithms.

Abstract: Social media can directly support disease management by creating online spaces where patients can interact with clinicians, and share experiences with other patients. Nevertheless, much more work remains to be carried out for providing and sharing an optimized information content. In this paper we formulate, from a theoretical perspective, an optimization problem aimed to encourage the creation of a sub-network of patients which, being recently diagnosed, wish to deepen their knowledge about their pathologies with some other patients, whose clinical profile turn to be similar, and have already been followed up within specific, even alternative, care centers. We will focus on the hardness of the proposed problem and provide a Genetic Algorithm (GA-based) approach to seek faster approximated solutions.

1 INTRODUCTION

The participatory, interactive nature of social media platforms allows for information to be generated and shared in a viral fashion, and provide new mechanisms to foster engagement and partnership with users and patients, to change their behaviors, and to fight against unhealthy lifestyles.

Due to their possible implications in public health, a growing number of scientists suggests to incorporate social media in health promotion and health care programs (Burke-Garcia and Scally, 2014). Social media can directly support disease management by creating online spaces where patients can interact with clinicians, and share experiences with other patients (Coiera, 2013; Santoro et al., 2015). For example, cancer patients use Twitter to discuss treatments and provide psychological support (Tsuya et al., 2014), and online engagement seems to correlate with lower levels of self reported stress and depression (Beaudoin and Tao, 2008).

Similarly, wellness programs frequently incorporate social media to create a sense of community (Zoppis et al., 2016), group people around shared goals, and offer social and emotional support. A trial reported that adding on line community features to an Internet-mediated wellness and walking program improves adherence, and did reduce participant attrition (Richardson et al., 2010).

Nevertheless, much more work remains to be carried out for sharing targeted and optimized information content. How can we optimize a procedure which is able to facilitate the encounter between patients who want to deepen or share experiences about treatments, care points, and specialists? How to correlate, for example, similar clinical profiles, while inducing networks of medical stuff, and treated patients which offer their availability to share experiences or suggestions? These are exactly the questions we try to answer in this paper.

It is clear that a proper handling of data is fundamental in order to convert available social spaces into useful sub-networks that leads to particular induced communities. Here, we focus on the problem of creating a space of individuals and care centers, by considering the case where recently diagnosed patients could be interested to meet some other patients (experience), for sharing information on their own disease or about the suggested (or available) care center. In this situation, it would be useful, for example, to encourage the diagnosed subjects to socialize, and confront with the experience of other patients of similar clinical profile, who have been already followed up within the same (or even alternative) proposed care

Zoppis I., Dondi R., Santoro E., Castelnuovo G., Sicurello F. and Mauri G.
Optimizing Social Interaction - A Computational Approach to Support Patient Engagement.
DOI: 10.5220/0006730606510657
In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (HEALTHINF 2018)*, pages 651-657
ISBN: 978-989-758-281-3

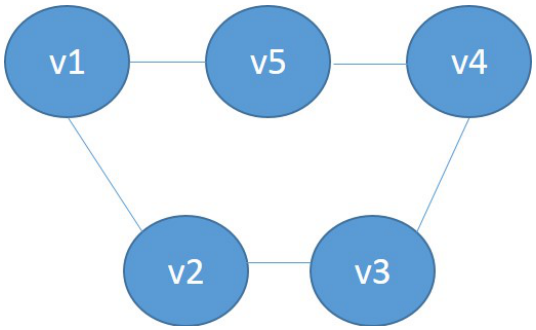

Figure 1: An example of a 2-club consisting of 5 vertices, such that each subgraph of four vertices is not a 2-club. Assume that we remove vertex v_5; then the graph induced by $\{v_1, v_2, v_3, v_4\}$ is not a 2-club ($d_{G'}(v_1, v_4) = 3$).

point.

In the next sections, we first introduce the main theoretical aspects, focusing on the computational hardness of the considered problem (Sections 3, 2.1, and 2.2). Then, we discuss a GA-based approach to seek faster approximated results to our questions[1] (Section 3). Finally, after reporting the numerical experiments on simulated data, we conclude the paper (Section 5) by describing future directions of this research.

2 MAIN THEORETICAL CONCEPTS

Suppose we wish to model the situation where recently diagnosed patients (shortly reported as RD patients) are fostered to deepen the knowledge of their diseases with some other patient experience (say, engaged patients or shortly, ED), or they need more information concerning the suggested (or even alternative) health-care centers (HC). In this case, RD patients could benefit from the social interaction with similarly profiled (ED) patients, which have already been followed up by specific HC centers. To this aim, it should be useful for a social platform to encourage, and optimize, the creation of a sub-network from the available social data.

From a theoretical perspective, a network is most commonly modeled using a graph, e.g., (Bollobás, 1998)) which represents relationships between objects, V (vertices), through a set of edges, E. In this way, our goal can be formulated by maximizing, within a defined graph, a *cohesive* sub-graph (i.e., by seeking the largest cohesive sub-graph) with particular properties. Such structures (i.e., cohesive or dense sub-networks) have been widely applied in several contexts. In computational biology, e.g., dense sub-graphs are sought in protein interaction networks, as they are considered related to protein complexes (Bader and Hogue, 2003; Spirin and Mirny, 2003), and, in gene networks, dense sub-graphs are applied to detect relevant co-expression clusters (Sharan and Shamir, 2000).

A classical approach to compute dense sub-graphs is the identification of cliques, i.e., complete sub-graphs induced by a set of vertices which are all pairwise connected by an edge. This definition of dense sub-graph is often too stringent for particular need. Indeed, some pair of elements may not be directly connected in a dense sub-graph, for example due to missing data that produce a dense sub-graph which is not, currently, a clique. Alternative definitions of cohesive subgraphs have been introduced, for example by relaxing some constraint of the clique definition, leading to the concept of *relaxed clique* (Komusiewicz, 2016).

In this paper, we focus on relaxing the distance between vertices. In a clique distinct vertices are at distance of 1, in our case, vertices have to be at distance of at most $s = 2$. A sub-graph where all the vertices are at distance of at most 2 is called a 2-*club* (or, more in general, *s-club* for different values of s). When $s = 1$, a 1-club is exactly a clique. In Fig. 1 is represented a 2-club consisting of 5 vertices. 2-*clubs* have been extensively applied to social networks analysis (Mokken, 1979; Alba, 1973; Laan et al., 2016; Mokken et al., 2016), and biological network analysis, e.g., protein-protein interaction networks, (Pasupuleti, 2008).

2.1 Problem Formulation

Consider a graph $G = (V, E)$, and a subset $V' \subseteq V$. We denote by $G[V']$ the subgraph of G induced by V'. Formally $G[V'] = (V', E')$, where

$$E' = \{\{u, v\} : u, v \in V' \wedge \{u, v\} \in E\}.$$

Given a set $V' \subseteq V$, we say that V' *induces* the graph $G[V']$ [2]. The *distance* $d_G(u, v)$ between two vertices u, v of G, is the length of a shortest path in G which has u and v as endpoints. The *diameter* of a graph $G = (V, E)$ is $\max_{u,v \in V} d_G(u, v)$, i.e., the maximum distance between any two vertices of V. In other words, a 2-club in a graph $G = (V, E)$ is a sub-graph $G[W]$, with $W \subseteq V$, that has diameter of at most 2. We

[1] See (Mitchell, 1996) for details on genetic algorithms)

[2] Notice that all the graphs we consider are undirected.

will formulate the problem using 2-clubs whose shortest path connecting RD patients with specialist centers/staff (e.g., care center, hospital, or clinical staff) has to "transit" through, at least one EP patient who has already been followed up by the considered specialist "point".

Formally, for any pair, (d,h), composed by the recently diagnosed patient, d, and, e.g., the health care center, h, the social platform should suggest for the patient d, to compare (even to meet) with an identified (available) patient x's experience. More specifically, we are currently seeking (within the input "social graph") a 2-Club, $G[D \cup X \cup H]$, where D, X and H represent the sets of recently diagnosed patients, experienced patients, and care point centers. respectively. Please note that, when such a structure (i.e., a maximum size 2-clubs) exists, within the identified starting social space, then for any pair of vertices, it must exist at least one simple path of length 2, i.e., a path composed by a triple of vertices. This, in turn, will be also true for any pair, (d,h) where $d \in D, h \in H$. Indeed, our goal will be to find the largest-size 2-clubs which has the further property of providing, for any pair (d,h), a shortest path characterized by the triple of vertices $(d,x,h) \in (D \times X \times H)$. In this case, the set of edges, modeling the starting social network, will be defined as follow.

- Edges between similar profiled patients.
- Edges expressing the fact that an experienced patient x, has already been properly followed up from specialists or care centers, h. In this case the edges in $X \times H$ will be constructed by knowing both the clinical history of each (experienced) patient, x, and the clinical staff or hospital h which has already properly followed up the patients, x.
- Edges between two vertices $h_1, h_2 \in H$, for example because two care centers are similar (have similar services or are part of the same institution).

In this situation, the simple path given by the triple of vertices $(d,x,h) \in (D \times X \times H)$ in the 2-club G would suggest for patient $d \in D$ to contact the patient, $x \in X$, about the health care center (or specialist staff), $h \in H$. For sake of clarity, before defining computationally the problem, we refer to any pair of vertices, $(d,h) \in D \times H$ (such that the minimum path connecting d to h is given by the vertex sequence (d,x,h), for any $x \in X$), as a "feasible pair". Considering the above discussion, we can define the following variant of the 2-clubs maximization problem

Problem 1. *Maximum 2-Club (Max-2-club)*
Input: *a graph $G = (D \cup X \cup H, R \cup F)$.*
Output: *a set $V' \subseteq D \cup X \cup H$ such that $G[V']$ is a 2-club having maximum size, and for each pair of vertices $(d,h) \in D \times H$ in $G[V']$, a minimum path connecting d to h is given by the vertex sequence (d,x,h) for some $x \in X$ (i.e., (d,h) is feasible).*

2.2 Computational Hardness

The complexity of the problem of Maximum s-club has been extensively studied in literature, and unfortunately it turns to be NP-hard for each $s \geq 1$ (Bourjolly et al., 2002); Maximum s-Club is NP-hard even if the input graph has diameter $s+1$, for each $s \geq 1$ (Balasundaram et al., 2005). The same property holds for our variant of Maximum 2-club. Indeed, the computation of a 2-club of maximum size containing a specific vertex v is also NP-hard. By defining $D = \{v\}$, $X = N(v)$ and H the remaining set of vertices, it follows that the "feasibility" property holds.

Given an input graph $G = (V,E)$, Maximum s-club is not approximable within factor $|V|^{1/2-\varepsilon}$, for any $\varepsilon > 0$ and $s \geq 2$ (Asahiro et al., 2010). On the positive side, polynomial-time approximation algorithms (Asahiro et al., 2010) have been given, with factor $|V|^{1/2}$ for every even $s \geq 2$, and factor $|V|^{2/3}$ for every odd $s \geq 3$. The parameterized complexity of Maximum s-Club has also been studied, leading to fixed-parameter algorithms (Schäfer et al., 2012; Komusiewicz and Sorge, 2015; Chang et al., 2013). Maximum 2-Club has been considered also for specific graph classes (Hartung et al., 2015; Golovach et al., 2014).

3 A GENETIC ALGORITHM

The complexity of the problems introduced so far make optimization potentially impracticable. For this reason, we designed a Genetic Algorithm (GA) to seek faster approximation solutions see, e.g., (Mitchell, 1996) for details.

In particular, given an input graph $G = (V,E)$, the proposed GA represents a solution (a subset $V' \subseteq V$ such that $G[V']$ is a 2-club of G, with the property discussed above) as a binary chromosome c of size $n = |V|$, whose ith component is defined as follows: $c[i] = 1$, for all $v_i \in V'$, else $c[i] = 0$. During the offspring generation, chromosomes are interpreted as hypotheses of feasible solutions, or they can even represent unfeasible solutions (e.g., s-club with $s > 2$, disconnected graphs, or "unfeasible pairs", as defined

above), which can evolve into feasible, due to mutation, cross-over, and selection. Moreover, hypotheses (i.e., chromosomes) are evaluated through the fitness function defined as follows

$$f(\text{diam}, n, m) = \begin{cases} n^2 + m^2 & \text{if } 0 \leq \text{diam} \leq 2 \ ; \\ \frac{1}{(\text{diam}^2 + \text{n}^2 + \text{m}^2)} & \text{if diam} > 2 \ , \end{cases} \tag{1}$$

where n, m, and diam are, respectively, the number of vertices of the induced subgraph, $G[D \cup X \cup H]$, the number of its feasible pairs, $(d,h) \in D \times H$, and its diameter. The goal of the fitness function described in Eq. 1 is to endorse new populations by promoting those chromosomes which represent subgraphs with high number of vertices, high number of feasible pairs, and diameter of length at most equal to 2. Specifically, for any fixed diameter diam ≤ 2 the fitness grow proportionally to the number of vertices and feasible pairs, thus promoting dense subgraphs. Instead, for any diam exceeding 2, the fitness decreases asymptotically, penalizing in this way, large subgraphs. Moreover, to allow a proper evolution (with regard to the Maximum 2-club problem), we defined the following standard operators.

- Mutation. Three type of mutations are considered.
  - Base Mutation. Similarly to the standard case, each individual from the current population at time i is modified with a given probability (see details in the experimental section). In this case, mutation flips a bit of the selected chromosome c, in such a way that the corresponding vertex is either removed (i.e., bit flipped to 0) or added (i.e., bit flipped to 1) to the solution induced by c. Note that, deleting or adding vertices may induce unfeasible sub-graphs, since the property of being a 2-club is not "hereditary". On the other hand, such modifications can introduce the chance to overcoming local minimum.

  - Non Standard Mutation 1. In this case mutation has the objective to correct hypotheses (i.e., chromosomes) consistently and parsimoniously. Since any chromosome, by representation, induce a sub-graph $G[V']$ of $G[V]$, which in turn may reflects feasible solutions, such hypotheses are verified using the following principle. Given a selected chromosome c, a vertex v' is (randomly) sampled from the set $V_+ = \{v_i : c[i] = 1\}$ and the minimum length of simple paths connecting every pair $(v_i, v'), v_i \in V_+$ is checked to be consistent with the chromosome representation, i.e., since each chromosome "speculates" a feasible 2-club, for such hypothesis to be true, there must be, at least, a simple path of size at most equal to 2 connecting any $v_i \in V_+$ with v'. If a negative feedback is observed after this verification, then the sampled vertex v' is flipped to 0.

  - Non Standard Mutation 2. This modification has the objective to increment (parsimoniously) the size of a solution. In this case, given a selected chromosome c a vertex v' is sampled from $V_- = \{v_j : c[j] = 0\}$ and the minimum length of simple paths connecting every pair (v_i, v') is checked to be consistent with the current representation of the chromosome c. In this case, we consider to extend the hypothesis represented by c, by adding v' to V_+ if minimum distances from v' to vertices of V_+ are not larger than 2.

- Cross-over. The following operations are provided.
  - Standard cross-over. Offspring is generated by copying and mixing parts of parents' chromosomes.

  - Logical AND between parents. This operation has the objective to provide an offspring consistent with the selected parents. For this, pairs of chromosomes are generated through logical AND operations between the ascendents.

  - Logical OR between parents. This operation has the objective to provide offspring extending parent hypotheses. Extension is given by realizing a logical OR operation between two selected parents.

- Elitist selection (or elitism). In order to guarantee that solution quality does not decrease from one generation to another (Baluja and Caruana, 1995), best hypotheses (high fitness values) are allowed to be part of a new offspring.

4 RESULTS

The genetic algorithm described in Sec. 3 was coded in R using the Genetic Algorithm package (Scrucca, 2013) downloadable at https://cran.r-project.org/web/packages/GA/index.html.

Results are given for synthetic data obtained by generating Erdos-Renyi (ER) random graphs $ER(n, p)$ with two free parameters: the number of vertices, n, of the input graph, and the probability, p, to

Table 1: Models (Erdos-Renyi). Input Diameter (InD), Output Diameter (OutD), Input Nodes (InN), Output Nodes (OutN), Output Feasible Pairs (OutP), Number of Unfeasible Diameters (UnD), Number of Unfeasible Pairs (UnDH), Best Fit (Fit), Ratio between input and output vertices (Rat.).

Model	InD	OutD	InN	OutN	OutP	UnD	UnDH	Fit	Rat.
ER(45,1/5)	3.2	2	45	13.4	16.6	0	0	466.4	3.36
ER(30,1/5)	4	2	30	10.4	8	0	0	189.6	2.88
ER(15,1/5)	5	2	15	5.2	5	0	0	36.6	2.88

Table 2: CPU time for Models in Tab. 1. CPU User Time (T_1), CPU System Time (T_2), CPU Elapsed Time (T_3) in seconds, Early stopping with no improvement (Run), Max Number of Generation (Iter).

Model	T1	T2	T3	Run	Iter
ER(45,1/5)	14658.28	10.132	14696.83	180	700
ER(30,1/5)	9965442	5.4	10003.26	180	700
ER(15,1/5)	5966.29	4.794	5988.46	180	700

Table 3: Models (Erdos-Renyi). In this case results are not averaged.

Model	InD	OutD	InN	OutN	OutP	UnD	UnDH	Fit	Rat.
ER(60,1/5)	3	2	60	15	14	0	0	421	4.00
ER(21,1/10)	8	2	21	8	3	0	0	45	2.63
ER(9,1/10)	4	2	9	6	6	0	0	72	1.50
ER(21,1/2)	3	2	21	14	18	0	0	520	1.50
ER(9,1/2)	9	2	9	6	4	0	0	52	1.50

Table 4: CPU time for Models in Tab. 3

Model	T1	T2	T3	Run	Iter
ER(60,1/5)	11169.94	8.69	11188.91	180	700
ER(21,1/10)	4971.67	3.91	4980.56	180	700
ER(9,1/10)	4981.36	3.06	5022.58	180	700
ER(21,1/2)	3832.23	3.49	3836.93	180	700
ER(9,1/2)	4707.55	3.1	4713.28	180	700

create edges between two vertices (Bollobas, 2001). Numerical experiments have the main objective to obtain, as reported above, feasible solutions which have the further property that, for any pair (d,h) of RD patient, d, and HC point, h, at least one ED patient is provided to make his experience available.

In all the experiments we applied a number, n, of vertices ranging in $\{9,15,21,30,45,60\}$, while the probability to create an edge is chosen in $\{1/2,1/5,1/10\}$. Moreover, we randomly labeled $n/3$ vertices as ED, $n/3$ as HC, and finally $n/3$ as AD vertices. Tables 1, 2, 3 and 4 report the performances of the system. First we executed the GAs iteratively by sampling the corresponding random model (i.e., 5 observations for each model), and the performance was averaged on the whole set of experiments (Tab. 1 and 2). The following attributes are reported.

- Input and output diameters. Graphs are represented as discussed in Sec. 3. The best GA solution is re-coded and the resulting diameter is reported (output diameter).
- Number of Input and output vertices.
- Number of (Final) Feasible Pairs. The resulting number of feasible pairs, (d,h), in the output (2-club) graph.
- Number of Unfeasible Solutions (Diameters). Total number of graphs (i.e., experiments) whose final diameters have dimension greater than 2 (after running the whole set of experiments).
- Number of Unfeasible Solutions (D-H Pairs). Total number of graphs (i.e., experiments) where at least one pair (d,h) does not provide one ED patient able to make his experience available.
- Fitness value. Fitness as described in Sec. 3.
- Ratio between the number of the input graph vertices and the number of vertices of the resulting 2-club.
- CPU User Time, CPU System Time, and CPU Elapsed Time in seconds.
- Early stopping for no improvement. The number of consecutive generations without improvement in the best fitness value before the GA is stopped.

- Max Number of Generation. The maximum number of iterations to run before the GA search is halted.

Another set of experiments was executed without repetitions by using different values of free parameters for each input graph. Tables 3, 4 reports the obtained performances. The following main considerations emerge from the results.

- All models effectively provide feasible 2-clubs with at least one experienced patient for each pair, (h,d), considered in the final solution.
- The models are able to find combinatorial structures which actually requires impractical computational time. This is the case of solutions given for input graph with high number of vertices (e.g., more of 40 vertices). In order to give an idea of the quality of the returned solutions, we have considered the ratio between the vertices of the input graphs and of the output graphs. Indeed, due to the computational hardness of the problem, we cannot compare the size of the subgraph returned by the GAs with the size of an optimal solution for the Maximum 2-club problem. Notice that the approximation complexity results for Maximum 2-club shows the problem is not approximable within factor $|V|^{1/2-\varepsilon}$, for each $\varepsilon > 0$ (Asahiro et al., 2010), thus the approximability is very hard to obtain. For this reason, we can say that, solutions which offer a ratio smaller than or close to two, are effectively compelling and interesting.

5 CONCLUSION

In this paper we focused on the problem of optimizing the creation of a sub-network of patients aimed to deepen the knowledge about the available care centers for their pathologies through the help of other "experienced" patients. We considered this problem from a computational point of view by defining a variant of the max 2-club problem.

The intrinsic complexity of the introduced formulation requires the use of heuristic algorithms to obtain feasible approximated solutions in a reasonable time. We showed that the proposed approach (GA-based) effectively provides empirical approximations able to find feasible structures (i.e., 2-clubs), which actually requires impractical computational time. In fact, while GAs optimization is not new in literature, a new design of these models is now needed to cope with the hardness of many computational problems which actually find new applications in many contexts (Dondi et al., 2017; Dondi et al., 2016).

From our results, it seems to emerge the possibility of extending this research using real data sets with larger instance's dimension. Moreover, a convergence analysis, and the use of tuning methods to optimize some free GA's parameter (e.g., probability values for choosing the available mutation or cross over operations, or even the use of alternative parameterized fitness functions) will be one of the future direction for this research.

Finally, it is important to emphasize that the framework described in this paper has to be considered, as discussed in Introduction, a tool to facilitate and promote the patient engagement. It is not clearly intended as an instrumentation to constrain the spontaneous nature of communication and interaction in a social network.

REFERENCES

Alba, R. D. (1973). A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology*, 3:113–126.

Asahiro, Y., Miyano, E., and Samizo, K. (2010). Approximating maximum diameter-bounded subgraphs. In *LATIN 2010: Theoretical Informatics, 9th Latin American Symposium, Oaxaca, Mexico, April 19-23, 2010. Proceedings*, pages 615–626.

Bader, G. D. and Hogue, C. W. V. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4:2.

Balasundaram, B., Butenko, S., and Trukhanov, S. (2005). Novel approaches for analyzing biological networks. *J. Comb. Optim.*, 10(1):23–39.

Baluja, S. and Caruana, R. (1995). Removing the genetics from the standard genetic algorithm. In Prieditis, A. and Russell, S. J., editors, *Machine Learning, Proceedings of the Twelfth International Conference on Machine Learning, Tahoe City, California, USA, July 9-12, 1995*, pages 38–46. Morgan Kaufmann.

Beaudoin, C. E. and Tao, C.-C. (2008). Modeling the impact of online cancer resources on supporters of cancer patients. *New Media & Society*, 10(2):321–344.

Bollobás, B. (1998). Random graphs. In *Modern Graph Theory*, pages 215–252. Springer.

Bollobas, B. (2001). *Random Graphs*. Cambridge University Press.

Bourjolly, J., Laporte, G., and Pesant, G. (2002). An exact algorithm for the maximum k-club problem in an undirected graph. *European Journal of Operational Research*, 138(1):21–28.

Burke-Garcia, A. and Scally, G. (2014). Trending now: future directions in digital media for the public health sector. *Journal of Public Health*, 36(4):527–534.

Chang, M., Hung, L., Lin, C., and Su, P. (2013). Finding large k-clubs in undirected graphs. *Computing*, 95(9):739–758.

Coiera, E. (2013). Social networks, social media, and social diseases. *BMJ: British Medical Journal (Online)*, 346.

Dondi, R., Mauri, G., and Zoppis, I. (2016). Clique editing to support case versus control discrimination. In *Intelligent Decision Technologies 2016*, pages 27–36. Springer.

Dondi, R., Mauri, G., and Zoppis, I. (2017). Orthology correction for gene tree reconstruction: Theoretical and experimental results. *Procedia Computer Science*, 108:1115–1124.

Golovach, P. A., Heggernes, P., Kratsch, D., and Rafiey, A. (2014). Finding clubs in graph classes. *Discrete Applied Mathematics*, 174:57–65.

Hartung, S., Komusiewicz, C., and Nichterlein, A. (2015). Parameterized algorithmics and computational experiments for finding 2-clubs. *J. Graph Algorithms Appl.*, 19(1):155–190.

Komusiewicz, C. (2016). Multivariate algorithmics for finding cohesive subnetworks. *Algorithms*, 9(1):21.

Komusiewicz, C. and Sorge, M. (2015). An algorithmic framework for fixed-cardinality optimization in sparse graphs applied to dense subgraph problems. *Discrete Applied Mathematics*, 193:145–161.

Laan, S., Marx, M., and Mokken, R. J. (2016). Close communities in social networks: boroughs and 2-clubs. *Social Netw. Analys. Mining*, 6(1):20:1–20:16.

Mitchell, M. (1996). *An introduction to genetic algorithms*. Complex adaptive systems. MIT press, Cambridge (Mass.).

Mokken, R. (1979). Cliques, clubs and clans. *Quality & Quantity: International Journal of Methodology*, 13(2):161–173.

Mokken, R. J., Heemskerk, E. M., and Laan, S. (2016). Close communication and 2-clubs in corporate networks: Europe 2010. *Social Netw. Analys. Mining*, 6(1):40:1–40:19.

Pasupuleti, S. (2008). Detection of protein complexes in protein interaction networks using n-clubs. In Marchiori, E. and Moore, J. H., editors, *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, 6th European Conference, EvoBIO 2008, Naples, Italy, March 26-28, 2008. Proceedings*, volume 4973 of *Lecture Notes in Computer Science*, pages 153–164. Springer.

Richardson, C. R., Buis, L. R., Janney, A. W., Goodrich, D. E., Sen, A., Hess, M. L., Mehari, K. S., Fortlage, L. A., Resnick, P. J., Zikmund-Fisher, B. J., et al. (2010). An online community improves adherence in an internet-mediated walking program. part 1: results of a randomized controlled trial. *Journal of medical Internet research*, 12(4).

Santoro, E., Castelnuovo, G., Zoppis, I., Mauri, G., and Sicurello, F. (2015). Social media and mobile applications in chronic disease prevention and management. *Frontiers in psychology*, 6.

Schäfer, A., Komusiewicz, C., Moser, H., and Niedermeier, R. (2012). Parameterized computational complexity of finding small-diameter subgraphs. *Optimization Letters*, 6(5):883–891.

Scrucca, L. (2013). GA: A package for genetic algorithms in R. *Journal of Statistical Software*, 53(4):1–37.

Sharan, R. and Shamir, R. (2000). Center CLICK: A clustering algorithm with applications to gene expression analysis. In *Proc. of the Eighth Int. Conf. on Int. Sys. for Mol. Biol., La Jolla / San Diego, CA, USA*, pages 307–316.

Spirin, V. and Mirny, L. A. (2003). Protein complexes and functional modules in molecular networks. *Proceedings of the National Academy of Sciences*, 100:12123–12–128.

Tsuya, A., Sugawara, Y., Tanaka, A., and Narimatsu, H. (2014). Do cancer patients tweet? examining the twitter use of cancer patients in japan. *Journal of medical Internet research*, 16(5).

Zoppis, I., Mauri, G., Sicurello, F., Santoro, E., Pietrabissa, G., and Castelnuovo, G. (2016). Diabesity: A study for mhealth integrated solutions. In *International Conference on Wireless Mobile Communication and Healthcare*, pages 195–199. Springer.