A few words on Correlation Clustering

Géraud Le Falher

April 14, 2016

The idea behind clustering is, given a set of objects, to gather similar ones into clusters so that all objects belonging to one cluster are similar to each other while being dissimilar to objects from all the other clusters.

One way to formalize this problem is to consider objects as the nodes of a graph, whose edges weight encode similarity. Namely, weights sign denote similarity or not, while their absolute value indicate the strengh of the relationship.

The objective can then be formulated in two way, either maximizing agreements (the number of $+$ edges inside clusters and $-$ edges across clusters) or minimizing disagreement (the number of $-$ edges within clusters and $+$ edges between clusters):

MaxAgree

$$\max_{\mathcal{S}} \sum_{(u,v)} c(w_{uv}) \left[w_{uv} > 0\right] \left[\mathcal{S}(u) = \mathcal{S}(v)\right] + c(w_{uv}) \left[w_{uv} < 0\right] \left[\mathcal{S}(u) \neq \mathcal{S}(v)\right] \tag{1}$$

MinDisAgree

$$\min_{\mathcal{S}} \sum_{(u,v)} c(w_{uv}) \left[w_{uv} < 0\right] \left[\mathcal{S}(u) = \mathcal{S}(v)\right] + c(w_{uv}) \left[w_{uv} > 0\right] \left[\mathcal{S}(u) \neq \mathcal{S}(v)\right] \tag{2}$$

where $\mathcal{S}$ is a clustering and $\mathcal{S}(u)$ is the cluster index of u and c is a cost function.

1 A bit of history

Ben-Dor *et al.* [1] consider an similar problem in the context of gene expression. They assume there exists a perfect clustering, and receive as input a complete similarity matrix, which is corrupted by measurement error. They give an $O(n^2(\log n)^c)$ algorithm that recover the planted disjoint cliques with high probability. Later, Shamir *et al.* [2] study the closely related Cluster editing problem: how few edges to add and remove in the input graph to turn it into a vertex-disjoint union of cliques?[1] They showed the problem is NP-complete, even if the number of clusters $p \geq 2$[2] is set beforehand and provide a

[1] A recent reference on exact parametrized decision version of this problem is [3].

[2] The reduction is from 3-exact 3 cover, see [2, Theorems 1, 2 and Corollary 1].

0.878 in weighted $p = 2$ case using a standard SDP relaxation. Independently, Bansal *et al.* [4] introduce problems (1) and (2) over complete graph under the name Correlation Clustering. They give a (large) constant approximation of MinDisAgree and a PTAS for MaxAgree.

Two surveys on the topic: [5, 6]

2 Hardness and approximation

Charikar *et al.* [7] show a 4-approximation on MinDisAgree for complete graph using a LP relaxation. On general graph, MinDisAgree also admits a LP relaxation in connection with the multicut problem with a $O(\log n)$ approximation rounding a SDP.

In addition to independently giving the same approximations as [7], Demaine *et al.* [8] show the equivalence between Correlation Clustering and the weighted multicut problem, which ask for the minimum weight set of edges whose removal in G disconnect the k pairs (s_i, t_i)

Charikar *et al.* [9, Section 4] give a $\Omega(\frac{1}{\log n})$ approximation of the MaxCorr problem, which is maximizing (1) - (2) and can be formulated as a quadratic programming problem solved in polynomial time.

Giotis *et al.* [10] give an more simple reduction from Graph Min Bisection When k, the number of clusters is fixed, they provide PTAS on complete graph for MaxAgree in $nk^{O(\epsilon^{-3}\log(\frac{k}{\epsilon}))}$ and MinDisAgree in $n^{O(\frac{9^k}{\epsilon^2})}\log(n)$. They also provide the following summary of results on general graph for fixed k:

k	2	≥ 3
MaxAgree	0.878 (improved to 0.884 by [11])	0.7666 [12]
MinDisAgree	$O(\sqrt{\log n})$ as it reduces to Min 2CNF Deletion for which Agarwal *et al.* [13] give such an approximation	this can be reduced from k-coloring, which for any $\epsilon > 0$ is NP-complete to approximate within $n^{1-\epsilon}$ [14]

Among related problems, Ailon *et al.* [17] also consider Correlation Clustering. On complete unweighted graph, they got a 3 approximation using CC-Pivot. On weighted complete graph, they give two versions, one that is greedy and a more sophisticated one that involved LP but perform better. 2.5 approximation with probability constraints ($w^+_{uv} + w^-_{uv} = 1$) and 2 approximation if we add triangular inequality ($\forall k\ w^-_{ij} \leq w^-_{ik} + w^-_{kj}$). Using more sophisticated rounding of the LP, Chawla *et al.* [15] improve these approximation to 2.06 and 1.5 respectively.

Let say weights are bounded and assume w.l.o.g. they lie in $[-1, +1]$. Bansal *et al.* [4, Section 7] describe a linear cost function where an edge of weight x incurs a cost $\frac{1-x}{2}$ when it's inside a cluster and $\frac{1+x}{2}$ when it's between two clusters. It is consistent with edges of integer weight $+1$, -1 and 0, the last ones incurring the same cost however we put them.

		k	MinDisAgree	MaxAgree
Complete	unweighted		2.06 [15]	PTAS in $n2^{poly(\frac{1}{\epsilon})}$ [10]
	weighted		1.5 (triangular inequality) [15]	?
General	unweighted		$O(\log n)$ [4], optimal under UCG conjecture [16]	0.7666 [12]
	unweighted	$k = 2$	$O(\sqrt{\log n})$ [13]	0.884 [11]
	weighted		$O(\log n)$ Demaine *et al.* [8]	?

Table 1: Best results on various problem.

Note that given a graph with weights bounded in absolute value by M, we can always transform them so to respect the probability constraints.

Weak social balance ensure that by forbidding triangles with exactly one negative edge, complete signed graph are perfectly clusterable[18]. Thus triangular inequality can be seen as a weighted relaxation of it, which explain why it leads to better approximation.

3 Applications

As stated in [8, Section 5]:

- finding optimal k
- adding constraints to existing problem
- clustering without distance (or with several conflicting distance), which can also be also be linked to metric learning.
- Visualization of signed social graph [19]
- image segmentation [20, 21] http://www.tandfonline.com/doi/abs/10.1080/10106049.2015.1110207
- duplicate detection, also called entity resolution [22]
- coreference resolution is solved by McCallum *et al.* [23, Section 2.3] using a undirected graphical model on which performing inference is equivalent to Correlation Clustering. Elsner *et al.* [24] compared various heuristics with a bound of the optimal solution obtained through SDP relaxation and show that best performing ones are within few percents of it, provided they are followed by a local search step. The same problem is tackeld by Chatel *et al.* [25], albeit with a different approach.

- blockmodel membership. For instance, in biology, Mason *et al.* [26] analyze a signed co-expression networks of genes involved in embryonic stem cells to find which genes are related to pluripotency or self-renewal.
- haplotype assembly [27]

Learning to Divide and Conquer for Online Multi-Target Tracking

4 Variations and related problems

recovery under noise Already considered in [4, Section 6] "classic minimum Multicut problem, for which the current state-of-the-art algorithm gives a $\Theta(\log n)$ factor approximation [GVY93]. This also implies that assuming the Unique Games Conjecture, we cannot obtain a constant factor approximation in the worst-case [KV05]."[16]

But maybe we can do better on average case, which motivate the study of semi random model, where real graphs are seen as being obtained from the controlled perturbation of a perfectly clusterable graph.

[28] "Our analysis makes three contribution. First, we define a model in which we derive finite-sample error bounds for correlation clustering. Second, we study the asymptotic behavior of correlation clustering with respect to the density of the graph and the scaling of cluster sizes. And finally, we propose a statistical test for evaluating the significance of a clustering."

[29]

"They consider a generalization of this model where there is an adversary. In their model, for each edge, with probability $(\frac{1}{2}+\epsilon)$ we do not flip the initial edge label, and with probability $(\frac{1}{2}-\epsilon)$ the adversary decides whether to flip the edge label or not. They give an algorithm that finds a clustering of cost at most $1+O(n-\frac{1}{6})$ times the cost of the optimal clustering, as long as $\epsilon \geq n-\frac{1}{3}$. However, these average-case models and algorithms deal only with the special case of complete graphs, with all edges having unit cost."[16].

Edge sign prediction [30, Section 6] using global matrix kernel. While the works mentioned above are combinatorial algorithms over arbitrary, undirected graph, we can also give additional semantic to edges in directed graph, such as friendship/foeship, trust/distrust, activator/inhibitor, secure/malicious interactions or high/low status. Assuming these graphs display some regularity like social balance, this can be used as bias for learning algorithms. Some heuristics are local. For instance, Leskovec *et al.* [31] build a logistic regression model based on degree feature and frequency of triangle in which each edge is involved. Chiang *et al.* [32] consider longer cycle but note that global approach are more efficient. In particular, inspired by the fact the complete, balanced adjacency matrix has k disjoint block, they give condition under which it can be recovered from the observed matrix through low rank matrix factorization.

Communities detection In the context of signed social network, clustering can also be seen as affecting each to one or more communities [33] based on random walks (that are likely to stay within one community) [33, 34, 35, 36, 37, 38, 39, 40]

Online & active setting Beside batch setting, one can also consider active [41, 42] and online [43] framework.

Mathieu *et al.* [44] give a greedy algorithm that upon vertex arrival creates a singleton cluster and then merge all pairs of clusters for which it increases the total number of agreements. For MinDisAgree, this is $O(n)$-competitive algorithm and they show such ratio is optimal by exhibiting an instance[3] on which any strategy ends up with $n-k$ disagreements whereas optimal cost is k. On the MaxAgree side, this greedy strategy result in a 0.5-competitive algorithm. If it is randomly mixed with a Dense variation, it raises up to $0.5+\eta$, still far from the demonstrated 0.834 upper bound.

[45, Section 5] present another active algorithm but it looks pretty involved so I haven't read it yet: "algorithms for MinDisAgree$[k]$ with sublinear query complexity, but the running time of this solutions is exponential in n".

Overlapping correlation clustering [46] in this setting, objects are allowed to belong to more than one cluster. Given a complete weighted graph, output a labelling function $\ell : V \to 2^{|V|}$ that minimize:

$$C_{occ}(V,\ell) = \sum_{u,v} |H(\ell(u),\ell(v)) - w_{uv}|$$

where H is a similarity function between sets of labels, such as Jaccard similarity or a binary indicator of non empty intersection. These problems are NP-hard and authors provide approximations algorithm and map-reduce experimental results.

consensus clustering the goal is to output a clustering which best summarize the given input clusterings. It's a restricted case of Correlation Clustering where the negative weights obey the triangular inequality. Gionis *et al.* [47] give a deterministic 3 approximation and describe a sampling approach suited to large dataset. It was later improved by Bonizzoni *et al.* [48], which show that the minimization version is APX-hard, even with only 3 candidates but give a PTAS with a $\frac{4}{3}$ approximation for the maximization problem. Experimental evaluations are conducted by Bertolacci *et al.* [49] and Filkov *et al.* [50].

non binary edge labelling Here, "positive" edges are categorical (or colored) and the goal is to form clusters mostly made up of edges of one color. As a generalisation of Correlation Clustering, it's NP-hard and Bonchi *et al.* [51] give a random approximation algorithm with a ratio bounded by max degree. They also present a fixed

[3] two positive cliques A and B joined by positive edges except between $a \in A$ and $\{b_1,\dots,b_k\} \in B$. Those nodes are given first and thus form a cluster which yield at least one disagreement every time one the $n-(k+1)$ remaining vertex is added.

k method and experiment on real world datasets. A linear programming-based algorithm by Anava *et al.* [52] achieves an improved approximation ratio of 4.

Bipartite Correlation Clustering First 11 approximation [53], then 4 approximation using LP [54] and improved to 3 even for K-partite graph [15]. If the solution is constrained to have k cluster and the graph is bipartite complete (or dense enough at least), there is a $(1-\delta)$ approximation running in time exponential in k and δ^{-1} but linear in $|V|$. The approximation can also be achieved with $O(\delta^{-1})$ [55].

Hypergraphs Kim *et al.* [21] show a LP relaxation on hypergraph and Ricatte *et al.* [56] describe a class of hypergraphs that can be reduced to signed graph.

Other approaches

spectral clustering and embedding [57]

parametrized cluster editing basically for instances where there only a few edges to edit per vertex, a polynomial time algorithm is presented in[58].

exact solution By casting the problem as a MaxSAT instance, one can take advantage of existing solvers to get an exact solution on small instances ($n \leq 1000$) [59]

probabilistic formulation looks like there is some work about that in the vision community (to segment pictures) [60]

4.1 Scalability

Distributed computation Chierichetti *et al.* [61] described a distributed approach with the following performance bound:

	# rounds	ratio
complete	$O(\log n)$	$3 + O(\epsilon)$
general	$O(\log n \log \Delta^+)$	$O(\log n)$

In the case of unit weight, this can be improved to $O(\log \log n)$ rounds and 3 approximation [62].

There are other approaches [6, Part III] one can look at.

A promising one is to perform Ailon algorithm using multiple cores [63].

One can also sample the data beforehand, as in [49].

Local search Bonchi *et al.* [64] study the problem of finding cluster index consistently given a single node and making some queries to the adjacency matrix of a general unweighted graph. They first describe a modification of Ailon QuickCluster giving a $4 \cdot OPT + \epsilon n^2$ approximation in $O(\frac{1}{\epsilon})$ time after $O(\frac{1}{\epsilon^2})$ time and queries preprocessing. Then there is an additive $OPT + \epsilon n^2$ additive approximation in $poly(\frac{1}{\epsilon})$ time after a $poly(\frac{1}{\epsilon})$ time and $2^{poly(\frac{1}{\epsilon})}$ queries preprocessing which seems more involved. Because

their approach is localized, it is trivially parallelized and is independent of the graph size. Therefore one can obtain a complete clustering in time linear with the number of node to predict.

5 Future directions

Ranging from theoretical to practical ones:

- improve bounds, for instance by designing ad-hoc algorithm for specific case
- improve scalability, for instance using distributed non negative matrix factorization, an active area of research [65–67]
- apply these new methods to real datasets (social network, NLP) and interpret domain specific results

References

[1] A Ben-Dor *et al.*, "Clustering gene expression patterns.", en, *J. Comput. Biol.*, vol. 6, no. 3-4, pp. 281–97, 1999. DOI: 10.1089/106652799318274 (cit. on p. 1).

[2] R. Shamir *et al.*, "Cluster graph modification problems", in *Graph-Theoretic Concepts Comput. Sci.* Vol. 2573, 2002, pp. 379–390. DOI: 10.1007/3-540-36379-3_33 (cit. on p. 1).

[3] F. V. Fomin *et al.*, "Tight bounds for parameterized complexity of cluster editing with a small number of clusters", *J. Comput. Syst. Sci.*, vol. 80, no. 7, pp. 1430–1447, 2014. DOI: 10.1016/j.jcss.2014.04.015 (cit. on p. 1).

[4] N. Bansal *et al.*, "Correlation clustering", *43rd Annu. IEEE Symp. Found. Comput. Sci. 2002. Proceedings.*, pp. 238–247, 2002. DOI: 10.1109/SFCS.2002.1181947 (cit. on pp. 2–4).

[5] H. Becker, "A survey of correlation clustering", Columbia University, Tech. Rep., 2005 (cit. on p. 2).

[6] F. Bonchi *et al.*, "Correlation clustering: from theory to practice", in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. data Min.*, 2014 (cit. on pp. 2, 6).

[7] M. Charikar *et al.*, "Clustering with qualitative information", in *44th Annu. IEEE Symp. Found. Comput. Sci.*, 2003, pp. 524–533. DOI: 10.1109/SFCS.2003.1238225 (cit. on p. 2).

[8] E. D. Demaine *et al.*, "Correlation clustering in general weighted graphs", *Theor. Comput. Sci.*, vol. 361, no. 2-3, pp. 172–187, 2006. DOI: 10.1016/j.tcs.2006.05.008 (cit. on pp. 2, 3).

[9] M. Charikar *et al.*, "Maximizing quadratic programs: extending grothendieck's inequality", in *45th Annu. IEEE Symp. Found. Comput. Sci.*, 2004, pp. 54–60. DOI: 10.1109/FOCS.2004.39 (cit. on p. 2).

[10] I. Giotis *et al.*, "Correlation clustering with a fixed number of clusters", pp. 1167–1176, 2006 (cit. on pp. 2, 3).

[11] P. Mitra *et al.*, "Approximation algorithm for correlation clustering", in *2009 1st Int. Conf. Networked Digit. Technol. NDT 2009*, 2009, pp. 140–145. DOI: 10.1109/NDT.2009.5272169 (cit. on pp. 2, 3).

[12] C. Swamy, "Correlation clustering: maximizing agreements via semidefinite programming", pp. 526–527, 2004 (cit. on pp. 2, 3).

[13] A. Agarwal *et al.*, "O(sqrt log n) approximation algorithms for min uncut, min 2cnf deletion, and directed cut problems", in *Proc. Thirty-seventh Annu. ACM Symp. Theory Comput.*, 2005, pp. 573–581. DOI: 10.1145/1060590.1060675 (cit. on pp. 2, 3).

[14] D. Zuckerman, "Linear degree extractors and the inapproximability of max clique and chromatic number", *Theory Comput.*, vol. 3, no. 6, pp. 103–128, 2007. DOI: 10.4086/toc.2007.v003a006 (cit. on p. 2).

[15] S. Chawla *et al.*, "Near optimal lp rounding algorithm for correlation clustering on complete and complete k-partite graphs", in *STOC'15*, 2014 (cit. on pp. 2, 3, 6).

[16] K. Makarychev *et al.*, "Algorithms for semi-random correlation clustering", 2014 (cit. on pp. 3, 4).

[17] N. Ailon *et al.*, "Aggregating inconsistent information", *J. ACM*, vol. 55, no. 5, pp. 1–27, 2008. DOI: 10.1145/1411509.1411513 (cit. on p. 2).

[18] J. A. Davis, "Clustering and structural balance in graphs.", *Hum. relations*, vol. 20, pp. 181–187, 1967. DOI: 10.1177/001872676702000206 (cit. on p. 3).

[19] E. W. D. Luca *et al.*, "Spectral analysis of signed graphs for clustering, prediction and visualization", in *Proc. 2010 SIAM Int. Conf. Data Min.* 2010, ch. 48, pp. 559–570. DOI: 10.1137/1.9781611972801.49 (cit. on p. 3).

[20] S Bagon *et al.*, "Large scale correlation clustering optimization", Tech. Rep., 2011, p. 9 (cit. on p. 3).

[21] S. Kim *et al.*, "Higher-order correlation clustering for image segmentation", in *Adv. Neural Inf. Process. Syst.*, 2011, pp. 1530–1538 (cit. on pp. 3, 6).

[22] O. Hassanzadeh *et al.*, "Framework for evaluating clustering algorithms in duplicate detection", *Proc. VLDB Endow.*, vol. 2, no. 1, pp. 1282–1293, 2009. DOI: 10.14778/1687627.1687771 (cit. on p. 3).

[23] A. McCallum *et al.*, "Conditional models of identity uncertainty with application to noun coreference", in *Adv. Neural Inf. Process. Syst. 17*, 2005, pp. 905–912 (cit. on p. 3).

[24] M. Elsner *et al.*, "Bounding and comparing methods for correlation clustering beyond ilp", pp. 19–27, 2009 (cit. on p. 3).

[25] D. Chatel *et al.*, "Fast gaussian pairwise constrained spectral clustering", in *ECML/PKDD - 7th Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discov. Databases*, 2014 (cit. on p. 3).

[26] M. J. Mason *et al.*, "Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells.", *BMC Genomics*, vol. 10, no. 1, p. 327, 2009. DOI: 10.1186/1471-2164-10-327 (cit. on p. 4).

[27] S. Das *et al.*, "Sdhap: haplotype assembly for diploids and polyploids via semi-definite programming.", *BMC Genomics*, vol. 16, no. 1, p. 260, 2015. DOI: 10.1186/s12864-015-1408-5 (cit. on p. 4).

[28] T. Joachims *et al.*, "Error bounds for correlation clustering", in *Proc. 22nd Int. Conf. Mach. Learn. - ICML '05*, 2005, pp. 385–392. DOI: 10.1145/1102351.1102400 (cit. on p. 4).

[29] C. Mathieu *et al.*, "Correlation clustering with noisy input", *21st Annu. ACM-SIAM Symp. Discret. Algorithms*, pp. 712–728, 2010 (cit. on p. 4).

[30] J. Kunegis *et al.*, "The slashdot zoo: mining a social network with negative edges", in *Proc. 18th Int. Conf. World wide web - WWW '09*, 2009, p. 741. DOI: 10.1145/1526709.1526809 (cit. on p. 4).

[31] J. Leskovec *et al.*, "Predicting positive and negative links in online social networks", in *Proc. 19th Int. Conf. World wide web - WWW '10*, 2010, p. 641. DOI: 10.1145/1772690.1772756 (cit. on p. 4).

[32] K.-y. Chiang *et al.*, "Prediction and clustering in signed networks: a local to global perspective", *J. Mach. Learn. Res.*, vol. 15, pp. 1177–1213, 2014 (cit. on p. 4).

[33] B. Yang *et al.*, "Community mining from signed social networks", *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 10, pp. 1333–1348, 2007. DOI: 10.1109/TKDE.2007.1061 (cit. on p. 5).

[34] V. A. Traag *et al.*, "Community detection in networks with positive and negative links", *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 80, 2009. DOI: 10.1103/PhysRevE.80.036115 (cit. on p. 5).

[35] P. Doreian *et al.*, "Partitioning signed social networks", *Soc. Networks*, vol. 31, no. 1, pp. 1–11, 2009. DOI: 10.1016/j.socnet.2008.08.001 (cit. on p. 5).

[36] P. Anchuri *et al.*, "Communities and balance in signed networks: a spectral approach", in *2012 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min.*, 2012, pp. 235–242. DOI: 10.1109/ASONAM.2012.48 (cit. on p. 5).

[37] A. Amelio *et al.*, "Community mining in signed networks", in *Proc. 2013 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min. - ASONAM '13*, 2013, pp. 95–99. DOI: 10.1145/2492517.2492641 (cit. on p. 5).

[38] Y. Li *et al.*, "A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks", *Soft Comput.*, vol. 18, no. 2, pp. 329–348, 2013. DOI: 10.1007/s00500-013-1060-4 (cit. on p. 5).

[39] Y Chen *et al.*, "Overlapping community detection in networks with positive and negative links", *J. Stat. Mech. Theory Exp.*, vol. 2014, no. 3, 2014. DOI: 10.1088/1742-5468/2014/03/P03021 (cit. on p. 5).

[40] J. Q. Jiang, "Stochastic blockmodel and exploratory analysis in signed networks", p. 12, 2015 (cit. on p. 5).

[41] N. Cesa-Bianchi *et al.*, "A correlation clustering approach to link classification in signed networks", *JMLR Workshop Conf. Proc.*, vol. 23, pp. 1–20, 2012 (cit. on p. 5).

[42] ——, "A linear time active learning algorithm for link classification", in *Adv. Neural Inf. Process. Syst. 25 (NIPS 2012)*, 2012, pp. 1–12 (cit. on p. 5).

[43] C. Gentile *et al.*, "Online similarity prediction of networked data from known and unknown graphs", *JMLR Workshop Conf. Proc.*, vol. 30, pp. 1–34, 2013 (cit. on p. 5).

[44] C. Mathieu *et al.*, "Online correlation clustering", in *27th Int. Symp. Theor. Asp. Comput. Sci. - STACS 2010*, Inria Nancy Grand Est & Loria, 2010, pp. 573–584 (cit. on p. 5).

[45] N. Ailon *et al.*, "Active learning using smooth relative regret approximations with applications", *J. Mach. Learn. Res.*, vol. 15, pp. 885–920, 2014 (cit. on p. 5).

[46] F. Bonchi *et al.*, "Overlapping correlation clustering", *Knowl. Inf. Syst.*, vol. 35, no. 1, pp. 1–32, 2012. DOI: 10.1007/s10115-012-0522-9 (cit. on p. 5).

[47] A. Gionis *et al.*, "Clustering aggregation", *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, 4–es, 2007. DOI: 10.1145/1217299.1217303 (cit. on p. 5).

[48] P. Bonizzoni *et al.*, "On the approximation of correlation clustering and consensus clustering", *J. Comput. Syst. Sci.*, vol. 74, no. 5, pp. 671–696, 2008. DOI: http://dx.doi.org/10.1016/j.jcss.2007.06.024 (cit. on p. 5).

[49] M. Bertolacci *et al.*, "Are approximation algorithms for consensus clustering worthwhile?", in *Proc. 2007 SIAM Int. Conf. Data Min.* 2007, ch. 41, pp. 437–442. DOI: 10.1137/1.9781611972771.41 (cit. on pp. 5, 6).

[50] V. Filkov *et al.*, "Consensus clustering algorithms: comparison and refinement", in *Proc. Tenth Work. Algorithm Eng. Exp.* 2008, ch. 10, pp. 109–117. DOI: 10.1137/1.9781611972887.11 (cit. on p. 5).

[51] F. Bonchi *et al.*, "Chromatic correlation clustering", in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '12*, 2012, p. 1321. DOI: 10.1145/2339530.2339735 (cit. on p. 5).

[52] Y. Anava *et al.*, "Improved theoretical and practical guarantees for chromatic correlation clustering", pp. 55–65, 2015 (cit. on p. 6).

[53] N. Amit, "The bicluster graph editing problem", M.Sc. Thesis, Tel Aviv University, 2004 (cit. on p. 6).

[54] N. Ailon *et al.*, "Improved approximation algorithms for bipartite correlation clustering", *SIAM J. Comput.*, vol. 41, no. 5, pp. 1110–1121, 2012. DOI: 10.1137/110848712 (cit. on p. 6).

[55] M. Asteris *et al.*, "Bipartite correlation clustering – maximizing agreements", in *AISTATS*, 2016 (cit. on p. 6).

[56] T. Ricatte *et al.*, "A spectral framework for a class of undirected hypergraphs", Research Report, 2013 (cit. on p. 6).

[57] Q Zheng *et al.*, "Spectral embedding of signed networks", in *Proc. 2015 SIAM Int. Conf. Data Min.* 2015, ch. 7, pp. 55–63. DOI: 10.1137/1.9781611974010.7 (cit. on p. 6).

[58] F. N. Abu-Khzam, "On the complexity of multi-parameterized cluster editing", 2015 (cit. on p. 6).

[59] J. Berg *et al.*, "Cost-optimal constrained correlation clustering via weighted partial maximum satisfiability", *Artif. Intell.*, 2015. DOI: 10.1016/j.artint.2015.07.001 (cit. on p. 6).

[60] J. H. Kappes *et al.*, "Multicuts and perturb & map for probabilistic graph clustering", 2016 (cit. on p. 6).

[61] F. Chierichetti *et al.*, "Correlation clustering in mapreduce", in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '14*, 2014, pp. 641–650. DOI: 10.1145/2623330.2623743 (cit. on p. 6).

[62] K. Ahn *et al.*, "Correlation clustering in data streams", in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 2237–2246 (cit. on p. 6).

[63] X. Pan *et al.*, "Scaling up correlation clustering through parallelism and concurrency control", in *NIPS Work. Discret. Comb. Probl. Mach. Learn.*, 2014 (cit. on p. 6).

[64] F. Bonchi *et al.*, "Local correlation clustering", 2013 (cit. on p. 6).

[65] S. Bhojanapalli *et al.*, "Tighter low-rank approximation via sampling the leveraged element", 2014 (cit. on p. 7).

[66] T. Hastie *et al.*, "Matrix completion and low-rank svd via fast alternating least squares", 2014 (cit. on p. 7).

[67] M. Udell *et al.*, "Generalized low rank models", 2014 (cit. on p. 7).