Clusterability in Signed Graphs

Selections from
A Mathematical Bibliography of Signed and Gain Graphs and Allied Areas

Electronic Journal of Combinatorics: Dynamic Surveys in Combinatorics
Preliminary Eighth Edition
2009 September 23
(Working copy as of this date.)

Compiled by
Thomas Zaslavsky
Department of Mathematical Sciences
Binghamton University (SUNY)
Binghamton, New York, U.S.A. 13902-6000

E-mail: zaslav@math.binghamton.edu

Annotations. I try to describe the relevant content in a consistent terminology and notation, in the language of signed graphs despite occasional clumsiness (hoping that this will suggest generalizations), and sometimes with my [bracketed] editorial comments. I sometimes try also to explain idiosyncratic terminology, in order to make it easier to read the original item. Several of the annotations incorporate open problems (of widely varying degrees of importance and difficulty).

I use these standard symbols:

Γ is a graph \((V, E)\) of order \(n = |V|\), undirected, possibly allowing loops and multiple edges. It is normally finite unless otherwise indicated.

Σ is a signed graph \((V, E, \sigma)\) of order \(n\). \(|\Sigma|\) is its underlying graph. \(E_+, E_-\) are the sets of positive and negative edges and \(\Sigma_+, \Sigma_-\) are the corresponding spanning subgraphs (unsigned).

\([\Sigma]\) is the switching class of \(\Sigma\).

\(A(\ )\) is the adjacency matrix.

\(\sim\) means that two signed or gain graphs are switching equivalent (with the same underlying graph).

\(\cong\) means that two signed or gain graphs are switching isomorphic (with isomorphic underlying graphs).

\(\cong\) denotes isomorphism.

\(l(\ )\) is the frustration index (= line index of imbalance).

Some standard terminology (much more will be found in the Glossary (Zaslavsky 1998c)):

circle: The graph of a simple closed path, or its edge set.

Adj Adjacency matrix, eigenvalues.

Alg Algorithms.

Bal Balance (mathematical), cobalance.

Clu Clusterability.

d Duality (graphs, matroids, or matrices).

Exp Expository.

Exr Interesting exercises (in an expository work).

Fr Frustration (imbalance); esp. frustration index (line index of imbalance).

Gen Generalization.

KG Signed complete graphs.

Phys Applications in physics (partial coverage).

PsS Psychological, sociological, and anthropological applications (partial coverage).

Ref Many references.

SD Signed digraphs: mathematical properties.

SG Signed graphs: mathematical properties.

VS Vertex-signed graphs (“marked graphs”); signed vertices and edges.

WD Weighted digraphs.

WG Weighted graphs.
Nir Ailon, Moses Charikar, and Alantha Newman


Nikhil Bansal, Avrim Blum, and Shuchi Chawla

Clusterability index $Q$ [minimum number of inconsistent edges; see Dorcian and Mrvar (1996a) for notation] in signed complete graphs is NP-hard. Polynomial-time algorithms for approximate optimal clustering: up to a constant factor from $Q$ (§3); probably within $1 - \varepsilon$ of $|E| - Q$ for any $\varepsilon$ (i.e., maximizing consistent edges within $1 - \varepsilon$) (§4). §3: A 2-clustering within $3Q_2$ (Thm. 2). A clustering within $cQ$ where $c \approx 20000$ (Thm. 13). §4: A clustering within $\varepsilon n^2$ of $|E| - Q$ with high probability but slow in terms of $1/\varepsilon$ (Thm. 15). Asymptotically faster in terms of $1/\varepsilon$ (Thm. 22). The $1 - \varepsilon$ factor results from the fact that $|E| - Q = \binom{n}{2} - Q > \frac{1}{2} \binom{n}{2}$ [so is not strong]. §6: “Random noise”. §7: “Extensions”, considers edge weights in $[-1, 1]$ (thus allowing incomplete graphs). Thm. 23: An unweighted approximation algorithm will also approximate this case, assuming “linear cost”: $e$ costs $(1 - w(e))/2$ if within a cluster and $(1 + w(e))/2$ if between clusters. Thm. 24: The problem for clustering that minimizes the total weight of + edges outside clusters and - edges within clusters (“minimizing disagreements”) is APX-hard. [Improved in Charikar, Guruswami, and Wirth (2003a, 2005a), Swamy (2004a). Generalized in Demaine et al. (2006a).] [Added: 22 Sept 2009.]

Vladimir Batagelj
See also P. Doreian and W. de Nooy.


M. Behzad and G. Chartrand
\[ \Lambda_{BC} \] Their line graph $\Lambda_{BC}(\Sigma)$ of a signed simple graph $\Sigma$ (not defined explicitly) is $L(|\Sigma|)$ with an edge negative when its two endpoints are negative edges in $\Sigma$. They “color” as in Cartwright and Harary (1968a) (i.e., clustering). Characterized: $\Sigma$ with colorable line graphs. Found: the fewest colors for
line graphs of signed trees, $K_n$, and $K_{r,s}$. [For a more sophisticated kind of line graph see Vijayakumar (various) and Zaslavsky (1984c, 20xxa).]

John Bramsen

Algorithmic ideas for estimating $l(\Sigma)$. Remarks on clusterability.

Dorwin Cartwright and Frank Harary

“Coloring” is clustering as in Davis (1967a). Thm. 1 adds a bit to Davis (1967a). Thm. 3: The clustering is unique $\iff$ all components of $\Sigma^+$ are adjacent.


Moses Charikar, Venkatesan Guruswami, and Anthony Wirth

Conference version of (2005a).

Gary Chartrand

[Reprinted (1985a).]


“Corrected reprint” of (1977a).

James A. Davis

$\Sigma$ is “clusterable” if its vertices can be partitioned so that each positive edge is within a part and each negative edge joins different parts. Thm.: $\Sigma$ is clusterable $\iff$ no circle has exactly one negative edge. [See Doreian and Mrvar (1996a)].

Survey of triad analysis in signed complete digraphs; cf. e.g. Davis and Leinhardt (1972a), Wasserman and Faust (1994a). [Added: 28 April 2009.]

(PsS, SD: Clu(Gen): Exp)

James A. Davis and Samuel Leinhardt

In “ranked clusterability” the vertices of a signed complete, symmetric digraph are divided into levels. The set of levels is totally ordered. A symmetric pair, \( \{ +vw, +wv \} \) or \( \{ -vw, -wv \} \), should be within a level. For an asymmetric pair, \( \{ +vw, -wv \} \), \( w \) should be at a higher level than \( v \). Analysis in relation to both randomly generated and observational data. [Added: 28 April 2009.]

(PsS, SD: Clu(Gen))

Erik D. Demaine, Dotan Emanuel, Amos Fiat, and Nicole Immorlica


(SG: WG: Clu: Alg)

Erik Demaine and Nicole Immorlica


(SG: WG: Clu: Alg)

Patrick Doreian

Reviews the development of balance and clustering theory for signed (di)graphs in social psychology, mainly Doreian and Mrvar (1996a), Doreian and Krackhardt (2001a), and especially Hummon and Doreian (2003a). The difference between Heider’s (1946a) and Cartwright and Harary’s (1956a) models, and the need to combine them. [Added: 24 April 2009.]

(PsS: Exp: SD, Bal, Clu, Alg)


Review of de Nooy, Mrvar, and Batagelj (2005a). (SG, SD: Bal, Clu, PsS)


Signed graphs \( \Sigma_1, \ldots \) (“multiple indicators”) may be approximations of a hidden signed graph \( \Sigma \). Goals: detect whether \( \Sigma \) exists, and find an optimal clustering of \( \Sigma \). Methods: (1) Examine the \( \Sigma_j \) for compatibility via statistical tests. (2) Estimate \( \Sigma \) by \( \sum_j \sigma_j \). (3) Applies the clusterability index and algorithm of Doreian and Mrvar (1996a). ((2) implies using weighted signed
Patrick Doreian, Vladimir Batagelj, and Anuška Ferligoj

Ch. 10: “Balance theory and blockmodeling signed networks”. Thm. (pp. 305–306; proof by Martin Everett): The sizes of the partitions of \( V \) that minimize the clustering index (Doreian and Mrvar 1996a) are consecutive integers.

Patrick Doreian, Roman Kapuscinski, David Krackhardt, and Janusz Szczy-pula


§ 3.3: “Results for group balance”. Describes results from analysis of data on a small (social) group, in terms of frustration index \( l \) and a clusterability index \( \min_{k>2} 2\hat{P}_{k,5} \) (slightly different from the index in Doreian and Mrvar (1996a)), finding both measures (but more so the latter) decreasing with time.

Patrick Doreian and Andrej Mrvar

They propose indices for clusterability (as in Davis (1967a)) that generalize the frustration index of \( \Sigma \). Fix \( k \geq 2 \) and \( \alpha \in [0,1] \). For a partition \( \pi \) of \( V \) into \( k \) “clusters”, they define \( P(\pi) := \alpha n_+ + (1-\alpha)n_- \), where \( n_+ := \) number of positive edges between clusters, \( n_- := \) number of negative edges within clusters, and \( 0 \leq \alpha \leq 1 \) is fixed. The first proposed measure is \( \hat{P}_k := \min_k P(\pi) \), minimized over \( k \)-partitions. A second suggestion is the “negation-minimal index of generalized imbalance” [i.e., of clusterability], the smallest number of edges whose negation (equivalently, deletion) makes \( \Sigma \) clusterable. [Call it the ‘clusterability index’ \( Q(\Sigma) \).] [Note that \( P(\pi) \) effectively generalizes the Potts Hamiltonian as given by Welsh (1993a). Question. Does \( P(\pi) \) fit into an interesting generalized Potts model?] [\( P(\pi) \) also resembles the Potts Hamiltonian in Fischer and Hertz (1991a) (q.v. for a related research question).]

They employ a local optimization algorithm to evaluate \( P_{k,\alpha} \) and find an optimal partition: random descent from partition to neighboring partition, where \( \pi \) and \( \pi' \) are neighbors if they differ by transfer of one vertex or exchange of two vertices between two clusters. This was found to work well if repeated many times. [A minimizing partition into at most \( k \) clusters is equivalent to a ground state of the \( k \)-spin Potts model in the form given by Welsh (1993a), but not quite in that of Fischer and Hertz (1991a).]

Terminology: \( P(\pi) \) is called the “criterion function”. [More explicitly, one might call \( Q(\Sigma, \pi; \alpha) := 2P(\pi) \) the ‘\( \alpha \)-weighted clusterability index of \( \pi \),
so the clusterability index $Q(\Sigma) = \min_{\pi} Q(\Sigma, \pi; .5)$; and call $Q_k := 2P_k$ the ‘$k$-clusterability index’ of $\Sigma$. Clusterability is “$k$-balance” or “generalized balance”. Clusters are “plus-sets”. Signed digraphs are employed in the notation but direction is ignored.

[Further developments in Doreian et al. (various), Hummon and Doreian (2003a), Bansal et al. (2004a), Demaine et al. (2006a), Mrvar and Doreian (2009a).]

The data in Doreian (2008a), and common sense, suggest that clusters should be allowed to overlap. This is an unplumbed direction.

Added: 22 Sept 2009. (SD: sg: Bal, Clu: Fr(Gen), Alg, PsS)


Similar to (1996a). Some lesser theoretical detail; some new examples. The $k$-clusterability index $P_{k, \alpha}$ (1996a) is compared for different values of $k$, seeking the minimum. [But for which value(s) of $\alpha$ is not stated.] Interesting observation: optimal values of $k$ were small. It is said that positive edges between parts are far more acceptable socially than negative edges within parts [thus, in the criterion function $\alpha$ should be rather near 1].

(Added: 7 Feb. 2009.) (SD: sg: Bal, Clu: Fr(Gen), Alg, PsS)


Generalizes the ideas of (1996a) (q.v.). Given: A signed digraph $(\vec{\Gamma}, \sigma)$; a “criterion function” $P(\pi, \rho) := \alpha n^+ + (1 - \alpha)n^-$, where $\pi := \{B_1, \ldots, B_k\}$ partitions $V$ into “clusters”, $\rho : \pi \times \pi \rightarrow \{+, -\}$, $0 \leq \alpha \leq 1$ is fixed, and $n^\varepsilon := \text{number of edges } \vec{B}_i \vec{B}_j \text{ with sign } \varepsilon$ for which $\rho(B_i, B_j) = -\varepsilon$ (over all $i, j$). Objective: $(\pi, \rho)$, or simply $k := |\pi|$, that minimizes $P(\pi, \rho)$. What is new and most interesting is $\rho$; also new is using the edge directions.

Call $(\vec{\Gamma}, \sigma)$ “sign clusterable” if $\exists (\pi, \rho)$ with $P(\pi, \rho) = 0$. Clusterability is sign clusterability with $\rho = \rho_+$, where $\rho_+(B_i, B_j) := +$ if $i = j$, $-\varepsilon$ if $i \neq j$. Let $P(k) := \min \{P(\pi, \rho) : |\pi| = k\}$. Thm. 4: $P(k)$ is monotonically decreasing. [Thus, there is always an optimum $\pi$ with singleton clusters. Why this does not vitiate the model is not addressed.] Thm. 5: If $(\vec{\Gamma}, \sigma)$ is sign clusterable, then $(\vec{\Gamma}, -\sigma)$ also is. If $(\vec{\Gamma}, \sigma)$ is clusterable, then $(\vec{\Gamma}, -\sigma)$ is not clusterable with the same $\pi$ [provided $E \neq \emptyset$]. If $(\vec{\Gamma}, \sigma)$ is sign clusterable with $\rho = -\rho_+$, then $(\vec{\Gamma}, -\sigma)$ is clusterable with the same $\pi$.

Relocation”: Shift one vertex, or exchange two vertices, between blocks so as to decrease $P$, as in (1996a). This is said (but not proved) to minimize $P$.

Refinements discussed: partially prespecified blocks; null blocks (without outgoing edges); criterion functions with special treatment of null blocks.

Applications to standard test examples of social psychology.

Dictionary: “balanced” = clusterable; “relaxed balanced” = sign clusterable; “$k$-balanced” = clusterable with $|\pi| = k$; “relaxed structural balance blockmodel” = this whole system. [Added: 7 Feb. 2009.]
D. Emanuel and A. Fiat
See also E. Demaine.

2003a Correlation clustering—Minimizing disagreements on arbitrary weighted graphs.

Joseph Fiksel

K.H. Fischer and J.A. Hertz
Sect. 3.7: “The Potts glass”. The Hamiltonian (without edge weights) is
\[ H = -\frac{1}{2} \sum \sigma (e_{ij}) (k \delta (s_i, s_j) - 1). \]
[It is not clear that the authors intend to permit negative edges. If they are allowed, \( H \) is rather like Doreian and Mrvar’s (1996a) \( P(\pi) \).
Question. Is there a worthwhile generalized signed and weighted Potts model with Hamiltonian that specializes both to this form of \( H \) and to \( P? \)] [Also cf. Welsh (1993a) on the Ashkin–Teller–Potts model.] (Phys: sg, clu: Exp)

Ioannis Giotis and Venkatesan Guruswami


Terry C. Gleason and Dorwin Cartwright
“Colorable” = clusterable. The adjacency matrices of \( \Sigma^+ \) and \( \Sigma^- \) are employed separately. The arithmetic is mostly “Boolean”, i.e., \( 1 + 1 = 0 \). A certain integral matrix \( T \) shows whether or not \( \Sigma \) is clusterable. [Added: 11 Nov. 2008.] (SG: Clu, Adj)

Norman P. Hummon and Patrick Doreian

Presents a model for evolution of balance and clusterability (as in Davis 1967a) of a signed digraph and explores it via computer simulations.

Definitions: Given a signed digraph \( \Sigma \) and a partition \( \pi \) of \( V \), define the ‘clusterability’ \( c(\Sigma, \pi) := (\# \text{ negative edges within blocks of } \pi) + (\# \text{ positive edges between blocks}) \). Define \( \pi(\Sigma) := \text{any } \pi \text{ that minimizes } c(\Sigma, \pi) \).
Define \( \Sigma(v_i) := \{ v_i \bar{v}_j \in \bar{E}(\Sigma) \} \) with signs. (\( \Sigma \) models relations in a social group \( V \). \( \Sigma_i \) is the graph of relations perceived by \( v_i \).)

Initial conditions: Fixed \( |V| \), fixed “contentiousness” \( p := \text{the probability that an initial edge is negative, a fixed “communication” rule, random } \Sigma^0\)
and, for each \( v_i \in V \), \( \vec{\Sigma}^0_i := \vec{\Sigma}^0 \). At time \( t + 1 \), \( \vec{\Sigma}^t_i(v_i) \) changes to \( \vec{\Sigma}^{t+1}_i(v_i) \) to minimize \( d(d(\vec{\Sigma}^{t+1}_i, \pi(\vec{\Sigma}^t))) \). Then \( \vec{\Sigma}^{t+1}_j(i) \) changes to \( \vec{\Sigma}^{t+1}_i(v_i) \) for some \( v_j \) (depending on \( \vec{\Sigma}_i \) and the communication rule).

Computer simulations examined the types of changes and emerging clusterability of \( \vec{\Sigma}_i \) or \( \vec{\Sigma}^t \) as \( t \) increases, under four different communication rules, random initial conditions with various \( p \), and \( |V| = 3, 5, 7, 10 \). The outcomes are highly suggestive (see §4; \( p \) seems influential). [Problem. Predict the outcomes in terms of initial conditions through a mathematical analysis.]

[Added: 26 April 2009.] (SD, sg: Bal, Clu: Alg)(PsS)

**John E. Hunter**


(SG: Bal, Clu)

**Takehiro Inohara**


Assumption: all \( \sigma(i, i) = + \). Thm. 3: A signed complete digraph is clusterable if and only if \( \sigma(i, j) = - \) or \( \sigma(j, k) = \sigma(i, k) \) for every triple \( \{i, j, k\} \) of vertices (not necessarily distinct). [The notation is unnecessarily complicated.] (SD: Clu, PsS)


**Eugene C. Johnsen and H. Gilman McCann**


Balance and clustering analyzed via triples rather than edges. [Possible because the digraph is complete. A later analysis via triples is in Doreian and Krackhardt (2001a).] (SD: Bal, Clu)

**Jérôme Kunegis, Andreas Lommatzsch, and Christian Bauckhage**


**T.A. McKee**


**Andrej Mrvar and Patrick Doreian**

§2, “Formalization of block-modeling signed two-mode data”: A signed two-mode network is a bipartite signed simple graph with color classes $V_1, V_2$. The objective is partitions $\pi_1, \pi_2$ of $V_1, V_2$ that minimize a “criterion function” $P := \alpha_1 + (1 - \alpha)i_1$; usually $\alpha = .5$. $k_1 := |\pi_1|$ and $k_2 := |\pi_2|$, or other restrictions, may be specified. Definitions: $\pi_i := \{V_{i1}, \ldots, V_{ik_i}\}$. A “block” is a nonvoid set $E(V_{i1}, V_{ij})$. Its sign is the sign of the majority of edges, + if a draw. $e$ is “consistent” with $(\pi_1, \pi_2)$ if it is in a block of sign $\sigma(e)$. $i_\varepsilon :=$ number of inconsistent edges of sign $\varepsilon$. [Added: 17 August 2009.] (SG: Clu, PsS)

Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj


§10.3: “Triadic analysis.” Types of balance and clusterability, with the triads (order-3 induced subgraphs) that do or do not occur in each. Table 16, p. 209. “Balance-theoretic models”, is a chart of 6 related models. §§10.7, 10.10: “Questions” and “Answers.” Some are on balance models. §10.9: “Further reading.” [Added: 28 April 2009.] (SG, SD, PsS: Bal, Clu, Alg: Exp)

Philippa Pattison


Edmund R. Peay

Proposes an index of nonclusterability for signed graphs and generalizes to edges weighted by a linearly ordered set. (SG, Gen: Clu: Fr(Gen))


See mainly §3: “Structural consistency.” (sd: Gen: Bal, Clu)


§3.1: “Signed graphs and the theory of structural balance.” Many topics are developed in the exercises. Exercise 4.2.7 (from Phillips (1967a)). (SG, SD: Bal, Alg, Adj, Clu, Fr, PsS: Exp, Exr)

Tadeusz Sozański

$\Sigma$ denotes a signed $K_n$. The “level of balance” (“indice du niveau d’équilibre”) $\rho(\Sigma) :=$ maximum order of a balanced subgraph. [Complement of the vertex deletion number.] Define distance $d(\Sigma_1, \Sigma_2) := |E_{1+} \triangle E_{2+}|$. Say
\(\Sigma\) is \(p\)-clusterable if \(\Sigma^+\) consists of \(p\) disjoint cliques [its “clusters”]. Thm. 1 evaluates the frustration index of a \(p\)-clusterable \(\Sigma\). Thm. 2 bounds \(l(\Sigma)\) in terms of \(n\) and \(\rho(\Sigma)\). A negation set \(U\) for \(\Sigma\) “conserves” a balanced induced subgraph if they are edge-disjoint; it is (strongly) conservative if it conserves some (resp., every) maximum-order balanced induced subgraph. Thm. 3: Every minimum negation set conserves every balanced induced subgraph of order > \(\frac{2}{3}n\). Thm. 4: A minimum negation set can be ordered so that, successively negating its edges one by one, \(\rho\) never decreases.

Chaitanya Swamy


Stanley Wasserman and Katherine Faust


D.J.A. Welsh [Dominic Welsh]


Includes very brief treatments of some appearances of signed graphs.

Sect. 4.4: “The Ashkin–Teller–Potts model”. This treatment of the Potts model has a different Hamiltonian from that of Fischer and Hertz (1991a). It does not seem that Welsh intends to admit edge signs but if they are allowed then the Hamiltonian (without edge weights) is

\[-\sum \sigma(e_{ij})(\delta(s_i, s_j) - 1).

Up to halving and a constant term, this is Doreian and Mrvar’s (1996a) clusterability measure \(P(\pi)\), with \(\alpha = .5\), of the vertex partition induced by
Bo Yang, William K. Cheung, and Jiming Liu

Given a (positively weighted) signed (di)graph, the authors provide an algorithm for an approximate clustering. Input: The graph and a length parameter $l$. Step 1: Construct transition probabilities $p_{ij} := [\sigma_{ij}w_{ij}]^+/d(v_i)$. Step 2: Apply the probabilities in a random walk of length $\leq l$; the matrix of $l$-step probabilities is $(p_{ij})^l$. Combine in a cluster the vertices that are detect reasonable clusters.

Also, a cut algorithm for approximate clustering. A cluster is $X \subset V$ such that the total net degree $d^\pm(\Sigma:X) \geq d^\pm(X,X^c)$ and $d^\pm(X^c,X) \leq d^\pm(\Sigma:X^c)$. [Added: 11 Feb. 2009.]