Balance and Clustering in Signed Graphs

Thomas Zaslavsky

Binghamton University (State University of New York)

C R Rao
Advanced Institute of
Mathematics, Statistics and Computer Science

26 July 2010

Outline

1. Signed Graphs
2. Balance
3. Frustration
4. Covering Radius of a Cycle Code
5. Psychology/Sociology: Social Tension
6. Physics: The Non-Ferromagnetic Ising Model of a Spin Glass
7. Dynamics
8. Clustering
9. Clusterability
10. Correlation Clustering: An Attempt at Organizing Knowledge
11. Bipartite Clusterability
12. Psychology/Sociology: Back to the Beginning

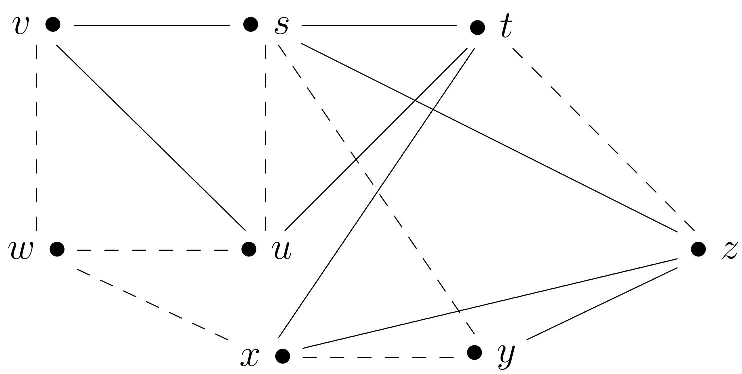
1. Signed Graphs

$\Sigma := (V, E, \sigma) = (|\Sigma|, \sigma)$ is a **signed graph**:

$|\Sigma| = (V, E)$ is the underlying graph: vertex set V, edge set E.
$\sigma : E \to \{+, -\}$ is the signature (sign function).

Positive subgraph: $\Sigma^+ := (V, E^+)$. Negative subgraph: $\Sigma^- := (V, E^-)$.

Σ_A



Switching:

Switching function $\zeta : V \to \{+, -\}$.
Switched signs: $\Sigma^\zeta := (|\Sigma|, \sigma^\zeta)$ defined by

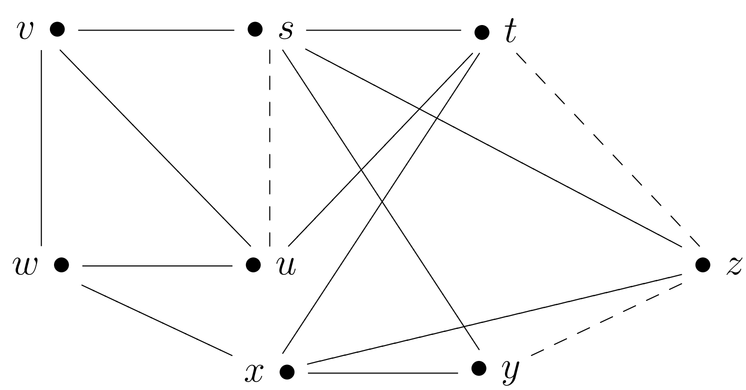
$$\sigma^\zeta(vw) := \zeta(v)\sigma(vw)\zeta(w).$$

Switching a set $X \subseteq V$: define $\Sigma^X := (|\Sigma|, \sigma^X)$ by

$$\sigma^X(vw) := \begin{cases} \sigma(vw) \text{ if } v, w \in X \text{ or } v, w \notin X, \\ -\sigma(vw) \text{ if } v \in X,\ w \notin X \text{ or } v \notin X,\ w \in X. \end{cases}$$

Switch $X = \{w, y\}$:

Σ_A^X



2. Balance

Sign of a circle C is $\sigma(C) :=$ product of edge signs.
Σ is balanced if every circle is positive.

Lemma 2.1. *Switching does not change the sign of any circle.*

Theorem 2.2. *The following statements are equivalent:*

(i) Σ *is balanced.*

(ii) (Harary's Balance Theorem) $V = V_1 \cup V_2$ *where* V_1, V_2 *are disjoint, and every positive edge is within* V_1 *or* V_2 *while every negative edge has one endpoint in each.*

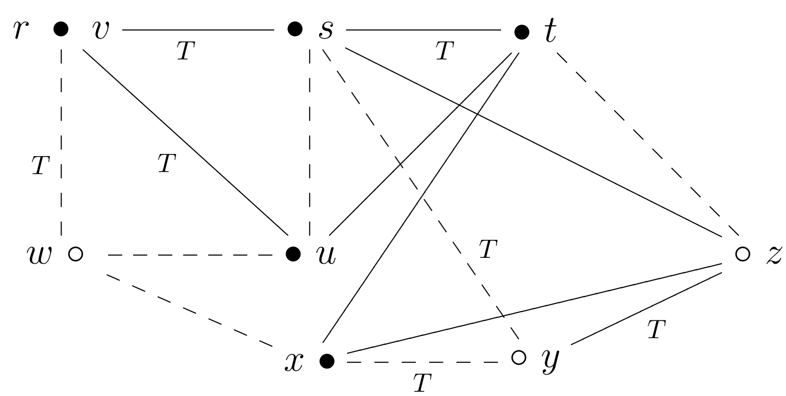
(iii) Σ *switches to an all-positive signature.*

Proof. (ii) $\implies$ (i): The negative edges form a cut, so every circle has an even number of negative edges.

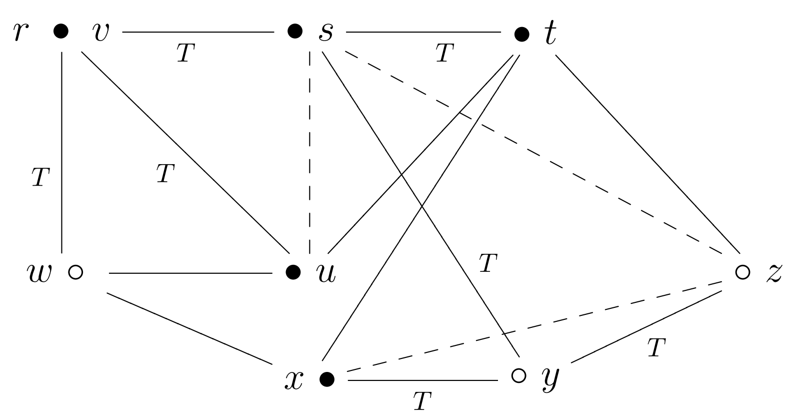
(iii) $\implies$ (ii): If Σ^X is all positive, let $V_1 = X$ and $V_2 = V \setminus X$.

(i) $\implies$ (iii): Choose a spanning tree T and a root r. Define $\zeta(v) := \sigma(T_{rv})$. In Σ^ζ, T is all positive, so there is a negative circle in $\Sigma \iff$ there is a negative edge in Σ^ζ. $\square$

Σ_A



Σ_A^ζ



Algorithm to Detect Balance:

(1) Choose T and r, and construct ζ.
(2) Switch to Σ^ζ.
(3) Check the sign of each edge, looking for negative edges.

Complexity: Fast.
(Let $n := |V|$.)

(1) Find T. Time n^2(?)
(2) Choose r. Time n^0.
(3) Construct ζ. Time n^1.
(4) Switch. Time n^1.
(5) Find a negative edge, if one exists. Time $O(n^2)$.

Total time: n^2.

The *First Mantra of Signed Graphs*: The basic fact is not the signs but the list of positive circles.

Theorem 2.3. *Given two signatures of the same graph, one can be switched to the other* $\iff$ *they have the same list of balanced circles.*

Corollary of the First Mantra: Signed graph theory is about switching classes, not individual signed graphs. (True mostly. Counterexample: Clusterability, in §§8, 9, 10.)

The *Second Mantra of Signed Graphs*: Everything that can be done for graphs can be done for signed graphs as well. (True very often! True mostly?)

3. Frustration

3.1. Measured by Frustration Index.

Frustration Index:
$l(\Sigma) :=$ least number of edges whose deletion makes Σ balanced.
A 'deletion set' $D \subseteq E$ satisfies: $\Sigma \setminus D$ is balanced.
A 'negation set' $N \subseteq E$ satisfies: Σ with the signs on N negated is balanced.

Proposition 3.1 (Traceable to Abelson & Rosenberg 1958)**.** *Frustration index is invariant under switching.* □

Theorem 3.2 (Harary)**.** *The least number of edges whose sign change makes Σ balanced = the least number whose deletion makes Σ balanced, $l(\Sigma)$.*

Proof. Any negation set is a deletion set. Any minimal deletion set is a negation set. Thus, minimal deletion sets and minimal negation sets are the same. □

Theorem 3.3. $l(\Sigma) = \min_\zeta |E^-(\Sigma^\zeta)|$*, the minimum over all switching functions.*

Proof. $l(\Sigma) \leq |E^-(\Sigma)| \implies l(\Sigma) \leq \min_\zeta |E^-(\Sigma^\zeta)|$.
Let D be a deletion set of size $l(\Sigma)$; then $\Sigma \setminus D$ is balanced, so $(\Sigma \setminus D)^\zeta$ is all positive for some switching function ζ. As $\Sigma^\zeta \setminus D$ is all positive, $D \subseteq E^-(\Sigma^\zeta)$. Thus, $l(\Sigma) = |D| \geq |E^-(\Sigma^\zeta)| \geq \min_\zeta |E^-(\Sigma^\zeta)|$. □

Lemma 3.4. *If $|E^-(\Sigma)| = l(\Sigma)$, then every vertex satisfies $d_{\Sigma^-}(v) \leq \frac{1}{2}d(v)$.*

Proof. If not, switch v, reducing the number of negative edges. □

Maximum Frustration:
For a graph Γ, $l_{\max}(\Gamma) := \max_\sigma l(\Gamma, \sigma)$.

Theorem 3.5 (Petersdorf 1966). $l(-K_n) = \lfloor (n-1)^2/4 \rfloor = l_{\max}(K_n)$.

Proof idea. For $l(-K_n)$ use the opposite of Harary's balance theorem: Find the biggest cut; the size of its complement is $l(-K_n)$.

Take $\Sigma = (K_n, \sigma)$, switched so $|E^-| = l(\Sigma)$. In Σ^-, the degrees are $d^-(v) \leq \lfloor \frac{n-1}{2} \rfloor$. This solves even n. For odd n, any two vertices with $d^-(v) = d^-(w) = \frac{n-1}{2}$ must be adjacent in Σ^-. Thus, there are $r \leq \frac{n-3}{2}$ of them; the other $n-r$ vertices have $d^-(x) \leq \frac{n-3}{2}$. Combining, $|E^-| \leq \frac{(n-1)^2}{4}$. □

There are estimates of $l_{\max}(\Gamma)$ for all graphs, bipartite graphs, etc., beginning with Akiyama, Avis, Era, & Chvátal 1981.)

No other significant infinite families are known.
Easy: $l_{\max}(C_n) = 1$, by $-C_n$ if n is odd but not if n is even.
Not hard: $l_{\max}(P) = 5 = l(-P)$, P = Petersen graph.

$l_{\max}(K_{r,s})$ is the obvious next candidate for solution after K_n. Very hard because $-K_{r,s}$ is balanced, thus, there is no candidate signature for maximum frustration. Good recent progress by Bowlin (2009) on $l_{\max}(K_{r,s})$ as a function of s, with r fixed; see §4, 'Covering Radius of a Cycle Code'.

Algorithm to Decide Frustration Index?

FRINDEX: 'Is $l(\Sigma) \leq k$?'

Proposition 3.6. *FRINDEX is NP-complete.*

Proof. FRINDEX for $-\Gamma$ is MAXCUT. □

Heuristics?
(1) Cutting plane methods (Grötschel et al.).
(2) Probabilistic methods. E.g.:
    Hill-climbing methods (local search with excursions).

Re (2): The structure of the state space? (See §6, 'Physics'.)

3.2. **Measured by Negative Triangles.**

Complete graphs only.

Triangle index: $$c_3(K_n, \sigma) := \frac{\text{number of negative triangles}}{\text{total number of triangles}}.$$

(Normalize to $[-1, 1]$. Then $c_3(-\Sigma) = -c_3(\Sigma)$.)

State Space:

$\Big\{(K_n, \sigma) : \sigma \in \{+, -\}^E\Big\}$, with adjacency when the signs differ on one edge.

Heuristic explorations by Antal et al., Marvel et al.

Jammed state:
No adjacent state has lower energy.

Questions: They exist. Where do they appear? What do they look like?

Valley:
Connected set of states (same energy) surrounded by higher-energy states.

Questions: Valleys that are not minimum energy? Do they exist? What are they like?

4. Covering Radius of a Cycle Code

Graph $\Gamma \mapsto$ cycle code $\mathcal{C}(\Gamma)$, the binary linear code generated by the circles.

Theorem 4.1 (Solé & Zaslavsky 1994)**.** *The covering radius of* $\mathcal{C}(\Gamma)$ *equals* $l_{\max}(\Gamma)$.

Best example:

$l_{\max}(K_{r,s})$ (§3.1, 'Maximum Frustration') is equivalent to finding the covering radius of the **Gale–Berlekamp code**, or equivalently, the badness of the worst case of the Gale–Berlekamp switching game.

Define:

$f_r(s) := l_{\max}(K_{r,s})$ as a function of s for fixed r:

Theorem 4.2 (Bowlin 2009, Theorem 4.22)**.** $f_r(s) = \dfrac{rs}{2}[1 - \delta_r(s)]$ *where*

$$\delta_r(s) \geq \frac{1}{s^{r-1}} \binom{r-1}{\lfloor \frac{r-1}{2} \rfloor},$$

with equality iff $2^{r-1} | s$.

Furthermore, $\delta_r(s)$ is eventually periodic with period 2^{r-1}, so it becomes very small compared with $f_r(s)$.

(Graham and Sloane 1985 had the bound, but without the condition for equality, and nonconstructively.)

5. Psychology/Sociology: Social Tension

(1) Heider (1946), 'Attitudes and cognitive organization.'
'P-O-X model': P, O are people, X is an object.

(2) Cartwright and Harary (1956), 'Structural balance: a generalization of Heider's theory.'
Social group with positive and negative relations.

(3) Davis (1967), 'Clustering and structural balance in graphs.'
More than two clusters with positive edges. (§8, 'Clustering'.)

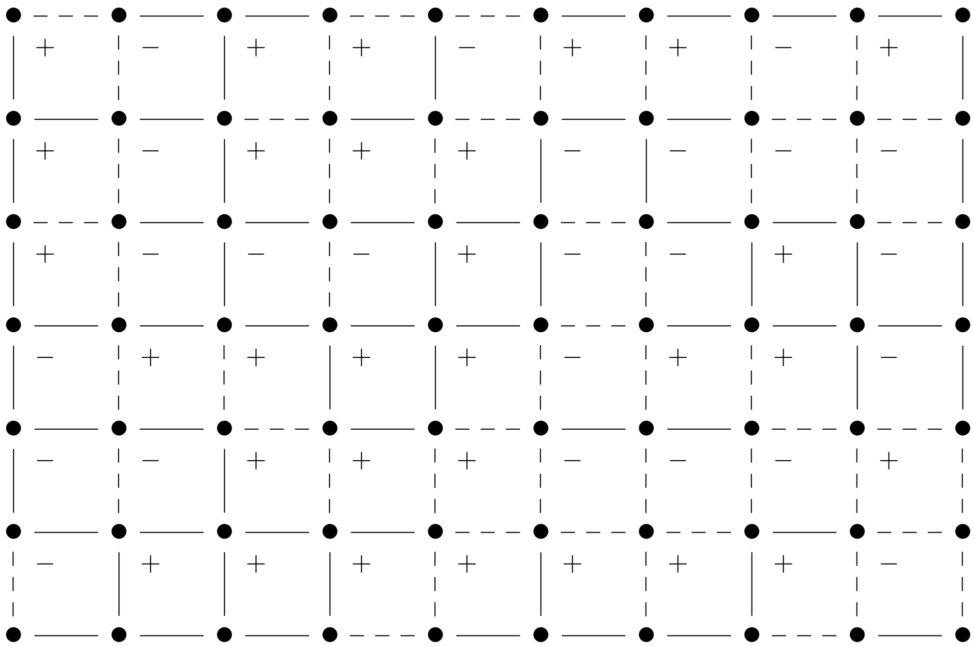
(4) Some experimental studies, of debatable outcome.

(5) Much theoretical interest in the context of social network theory, esp.

- Doreian et al., e.g. in Slovenia (Mrvar, Batagelj).
- Attempts to make the Cartwright-Harary theory more realistic.
- The bipartite model. (§11, 'Bipartite Clusterability'.)

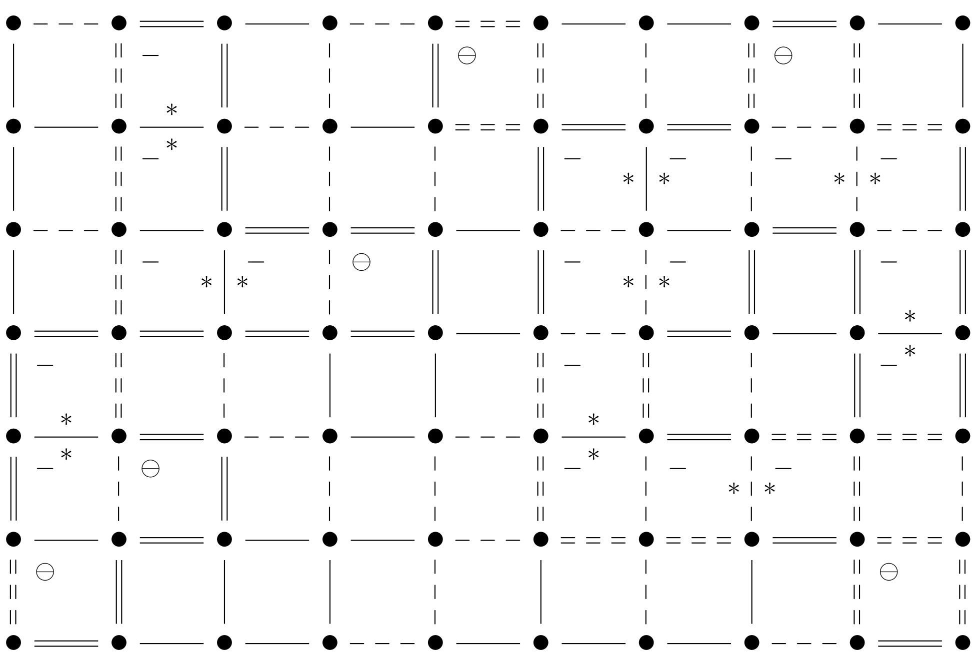
6. Physics: The Non-Ferromagnetic Ising Model of a Spin Glass

A square lattice representing a spin-glass:



Plaquettes: The little squares are satisfied (+) or frustrated (−).

Negative plaquettes form patterns.



Frustration index: 6 = number of unmatched frustrated plaquettes.

As shown here:

l(*planar* signed graph) is efficiently computable.
(Katai & Iwai, *J. Math. Psychology.*)

Same for *toroidal* signed graphs.
(Barahona.)

Not so for the 3-dimensional cubic lattice!

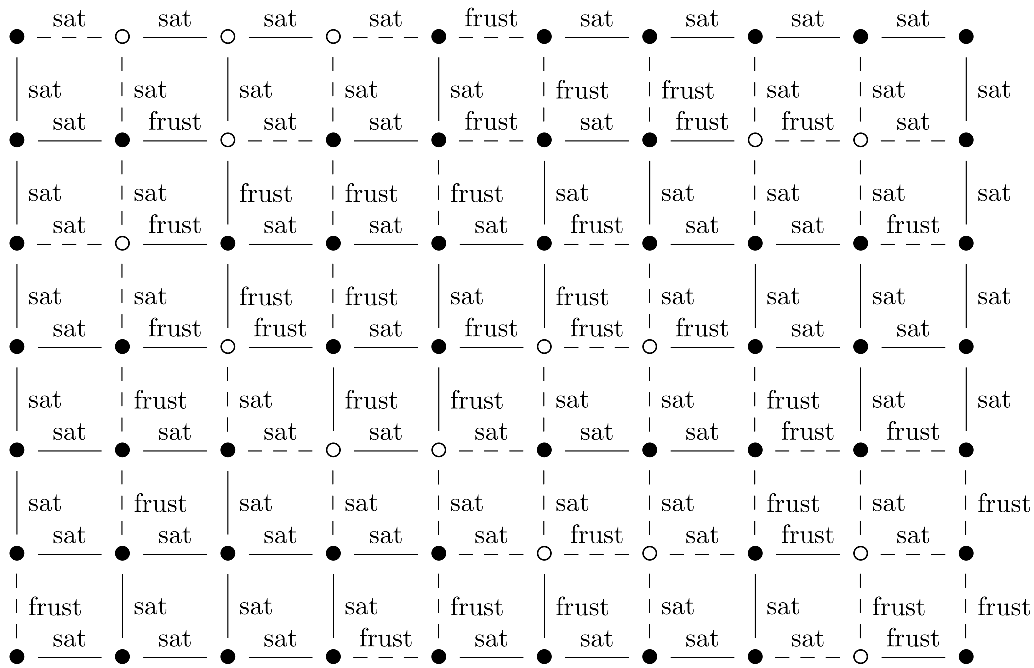
Fact: Every (finite) graph embeds in $\mathbb{Z}^3$.

Fact: $l(\Sigma)$ is NP-complete.

Conclusion: Bad news!

A state of the graph:

•: spin up (+) ○: spin down (−)
sat: satisfied edge frust: frustrated edge



Theorem 6.1. *A state that minimizes the number of frustrated edges is a switching function that reduces the number of negative edges to a minimum. Furthermore, the minimally frustrated edges are the same ones that become negative upon switching.*

Proof. Compare the definition of a frustrated edge with the signs resulting after switching. □

Hamiltonian: $H(\zeta) = 2|E| - |E^-(\Sigma^\zeta)|$.

Energy: $Ce^{-H(\zeta)}$, minimized at minimum $|E^-(\Sigma^\zeta)|$, i.e., when $|E^-(\Sigma^\zeta)| = l(\Sigma)$.

Thus, we are looking over the state space of Σ.
State space: $\{+,-\}^V$ with adjacency by one sign change. (The n-cube graph.)
Question: What is the energy landscape of the state space of Σ?

7. Dynamics

Make a state space and study the degree of balance.

I. States of a fixed signed graph; frustration index.
(See §3.1; also, §6, 'Physics'.)

II. Signatures of a fixed complete graph; triangle index.
(See §3.2, 'Triangle Index'.)

III. Signatures of a fixed graph; unknown index.

IV. Signatures of a fixed graph; clusterability indices.
(See §8, 'Clustering'; and §9, 'Clusterability'.)

Little is known, much is to be discovered.

8. Clustering

Σ is clusterable if $V = V_1 \cup V_2 \cup \cdots$ so all positive edges are within a cluster V_i and all negative edges are between clusters.

Σ is k-clusterable if it is clusterable with k clusters.

The obvious way to cluster: Find the components of Σ^+.

Theorem 8.1 (Davis 1967). *Σ is clusterable iff no circle has exactly one negative edge.*

Proof. Suppose Σ is clusterable. Consider C. If it has exactly one negative edge, its vertices must lie in a cluster, but its negative edge cannot.

Suppose the components of Σ^+ have vertex sets $V_1, \ldots, V_k$. If there is a negative edge e within a component $[V_i]$, there is a circle within $[V_i]$ that contains e. □

How do we decide whether Σ is k-clusterable? The graph obtained from $|\Sigma|$ by contracting each component of Σ^+ to a point is $|\Sigma|/E^+$. The chromatic number of Γ is $\chi(\Gamma)$ (∞ if Γ has a loop). $c(\Gamma)$ is the number of components of Γ.

Theorem 8.2. *Σ is $(\leq k)$-clusterable $\iff k \geq \chi(|\Sigma|/E^+)$. Σ is k-clusterable $\iff c(|\Sigma|) \geq k \geq \chi(|\Sigma|/E^+)$.*

Proof. A k-clustering combines into k groups components of Σ^+ that are not joined by negative edges of Σ. Color combined components the same and uncombined components differently; that is a k-coloring $|\Sigma|/E^+$ that uses all k colors, which is possible iff $c(|\Sigma|) \geq k \geq \chi(|\Sigma|/E^+)$. If Σ is not clusterable, $|\Sigma|/E^+$ has a loop so $\geq \chi(|\Sigma|/E^+) = \infty$. □

Theorem 8.3. *For $k = 2$, $\leq k$-clusterability is solvable in quadratic time. For $k > 2$, $\leq k$-clusterability and k-clusterability are NP-complete problems.*

Proof. 2-clusterability is balance, known to be quadratic. k-clusterability includes chromatic number, indeed $-\Gamma$ is $\leq k$-clusterable $\iff k \geq \chi(\Gamma)$ (a simple special case of Theorem 8.2). The question, 'Is Γ k-colorable?', is NP-complete for $k > 2$. The question 'Is Γ k-colorable using all k colors?', is easily equivalent. □

Not many signed graphs are clusterable, so we need to look deeper . . .

9. CLUSTERABILITY

k-clusterability index:
$Q_\Sigma(k)$:= the minimum number of bad edges in a k-clustering.

Theorem 9.1 (Doreian and Mrvar). $Q_\Sigma(k)$ *decreases to a minimum and then increases.*

Conclusion:
There is a contiguous range of cluster numbers that minimizes the number of bad edges.

Is it computable? (See §10, 'Correlation Clustering', next.)

Is it realistic?

Clusterability index:
$Q(\Sigma)$:= the minimum number of bad edges in any clustering
$= \min_k Q_\Sigma(k)$.

10. Correlation Clustering: An Attempt at Organizing Knowledge

Goal: Cluster related objects in the best possible way, automatically.
Example: Documents.
(Bansal, Blum, & Chawla 2004.)

Correlation clustering means maximizing correlation. This is identical to finding the clustering with the least clusterability index.

What's old or restrictive:

- Only K_n.
- Ignorant of social networks.

What's new:

- Specific algorithms and complexity considerations for:
  (1) Minimizing disagreements (bad edges).
  (2) Maximizing agreements (good edges).

Theorem 10.1 (Bansal, Blum, & Chawla 2004)**.** $Q(\Sigma)$ *is an NP-complete problem.*

Theorem 10.2 (Bansal, Blum, & Chawla 2004)**.** *A P-time algorithm to minimize the number of bad edges that gives a result no worse than* $20000\times$ *the actual number.*

Greatly improved by Charikar, Guruswami, & Wirth (2005).

Theorem 10.3 (Swamy 2004)**.** *An algorithm to maximize the number of good edges to within 0.7666, on a weighted signed graph.*

All generalized to weighted signed graphs by Demaine, Emanuel, Fiat, & Immorlica (2006).

11. Bipartite Clusterability

Σ is bipartite: $V = U \cup W$.

Clusters are within each part.
A (k_1, k_2)-biclustering is a partition of U into k_1 parts and of W into k_2 parts:

$$U = U_1 \cup U_2 \cup \cdots \cup U_{k_1}, \qquad W = W_1 \cup W_2 \cup W_{k_2}.$$

$Q_\Sigma(k_1, k_2) :=$ number of edges in $[U_i, W_j]$ that go against the majority.

Theorem 11.1 (Mrvar & Doreian 2009)**.** $Q_\Sigma(k_1, k_2)$ *is a weakly decreasing function.*

Proof. By splitting a part into two parts, one cannot increase the number of edges that go against the majority. □

Theorem 11.2 (Zaslavsky)**.** $Q_\Sigma(1, k_2), Q_\Sigma(k_1, 1) \geq l(\Sigma) \geq Q_\Sigma(2, 2)$.

Proof. Find $X \subseteq V$ such that Σ^X has $l(\Sigma)$ negative edges. Let

$$U_1 = U \cap X, \qquad W_1 = W \cap X,$$
$$U_2 = U \setminus X, \qquad W_2 = W \setminus X.$$

For $Q_\Sigma(1, k_2)$, the number of edges that go against the majority in each $[U, W_j]$ cannot be less than the number of negative edges in Σ^X.
For $Q_\Sigma(2, 2)$, the total number of edges in all $[U_i, W_j]$ that go against the majority, cannot be larger than the number of negative edges in Σ^X. □

Next step: Characterize the cases where $l(\Sigma) \geq Q_\Sigma(2, 2)$. This has been done but is too complicated and not important enough to describe here.

12. Psychology/Sociology: Back to the Beginning

Studies and inventions continue.

Clusterability:

Hummon and Doreian, Doreian and Mrvar:

- Testing k-clusterability index against small sample social groups.
- Looking for modifications or additions to the theory that give better results.

Biclusterability:

Mrvar and Doreian:
The two color classes are

U = set of people,
W = set of objects for which the people have favorable or unfavorable feelings.

Conclusions?

People are too complicated, but the models are mathematically interesting and have unexpected applications.

References

[1] Robert P. Abelson and Milton J. Rosenberg, Symbolic psycho-logic: a model of attitudinal cognition. *Behavioral Sci.* **3** (1958), 1–13.

[2] T. Antal, P.L. Krapivsky, and S. Redner, Dynamics of social balance on networks. *Phys. Rev. E* **72** (2005), 036121. MR 2006e:91124.

[3] Nikhil Bansal, Avrim Blum, and Shuchi Chawla, Correlation clustering. *Machine Learning* **56** (2004), no. 1–3, 89–113. Zbl 1089.68085.

[4] Francisco Barahona, Balancing signed toroidal graphs in polynomial-time. Unpublished manuscript, 1981.

[5] Francisco Barahona, On the computational complexity of Ising spin glass models. *J. Phys. A: Math. Gen.* **15** (1982), 3241–3253. MR 84c:82022.

[6] Francisco Barahona, On some applications of the Chinese Postman Problem. In: B. Korte, L. Lovász, H.J. Prömel, and A. Schrijver, eds., *Paths, Flows and VLSI-Layout*, pp. 1–16. Springer-Verlag, Berlin, 1990. MR 92b:90139. Zbl 732.90086.

[7] Dorwin Cartwright and Frank Harary, Structural balance: a generalization of Heider's theory. *Psychological Rev.* **63** (1956), 277–293. Reprinted in: Dorwin Cartwright and Alvin Zander, eds., *Group Dynamics: Research and Theory*, second ed., pp. 705–726. Harper and Row, New York, 1960. Also reprinted in: Samuel Leinhardt, ed., *Social Networks: A Developing Paradigm*, pp. 9–25. Academic Press, New York, 1977.

[8] Moses Charikar, Venkatesan Guruswami, and Anthony Wirth, Clustering with qualitative information. *J. Comput. System Sci.* **71** (2005), no. 3, 360–383. MR 2006f:68141. Zbl 1094.68075 .

[9] James A. Davis, Clustering and structural balance in graphs. *Human Relations* **20** (1967), 181–187. Reprinted in: Samuel Leinhardt, ed., *Social Networks: A Developing Paradigm*, pp. 27–33. Academic Press, New York, 1977.

[10] Erik D. Demaine, Dotan Emanuel, Amos Fiat, and Nicole Immorlica, Correlation clustering in general weighted graphs. *Theoretical Computer Sci.* **361** (2006), no. 2–3, 172–187. MR 2008e:68157. Zbl 1099.68074 .

[11] Patrick Doreian and Andrej Mrvar, A partitioning approach to structural balance. *Social Networks* **18** (1996), 149–168.

[12] Patrick Doreian and Andrej Mrvar, Partitioning signed social networks. *Social Networks* **31** (2009), no. 1, 1–11.

[13] Patrick Doreian, Vladimir Batagelj, and Anuška Ferligoj, *Generalized Blockmodeling.* Structural Analysis in the Social Sciences, No. 25. Cambridge Univ. Press, Cambridge, Eng., 2005.

[14] F. Harary, On the notion of balance of a signed graph. *Michigan Math. J.* **2** (1953–54), 143–146 and addendum preceding p. 1. MR 16, 733h. Zbl 056.42103.

[15] F. Harary, On the measurement of structural balance. *Behavioral Sci.* **4** (1959), 316–323. MR 22 #3696.

[16] Fritz Heider, Attitudes and cognitive organization. *J. Psychology* **21** (1946), 107–112.

[17] Osamu Katai and Sousuke Iwai, Studies on the balancing, the minimal balancing, and the minimum balancing processes for social groups with planar and nonplanar graph structures. *J. Math. Psychology* **18** (1978), 140–176. MR 83m:92072. Zbl 394.92027.

[18] Seth A. Marvel, Steven H. Strogatz,and Jon M. Kleinberg, Energy landscape of social balance. *Phys. Rev. Letters* **103** (2009), Article 198701.

[19] Andrej Mrvar and Patrick Doreian, Partitioning signed two-mode networks. *J. Math. Sociology* **33** (2009), no. 3, 196–221.

[20] Patrick Solé and Thomas Zaslavsky, A coding approach to signed graphs. *SIAM J. Discrete Math.* **7** (1994), 544–553. MR 95k:94041. Zbl 811.05034.

[21] Chaitanya Swamy, Correlation clustering: Maximizing agreements via semidefinite programming. In: *Proc. Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (New Orleans, 2004), pp. 526–527 (electronic). ACM, New York, and SIAM, Philadelphia, 2004. MR 2291092.

[22] Thomas Zaslavsky, A mathematical bibliography of signed and gain graphs and allied areas. *Electronic J. Combin.*, Dynamic Surveys in Combinatorics (1998), No. DS8 (electronic). MR 2000m:05001a. Zbl 898.05001. Current update at URL `http://www.math.binghamton.edu/zaslav/Bsg/`
Consult this for references not listed here.

[23] Thomas Zaslavsky, Frustration vs. clusterability in two-mode signed networks (signed bipartite graphs). Submitted.