Exploiting Social Network Structure for Person-to-Person Sentiment Analysis
(Supplementary Material to [WPLP14])

Robert West, Hristo S. Paskov, Jure Leskovec, Christopher Potts
Stanford University
west@cs.stanford.edu, hpaskov@stanford.edu, jure@cs.stanford.edu, cgpotts@stanford.edu

1 Exploiting symmetries (Sec. 3.2)

Our definition of a triangle $t = \{e_1, e_2, e_3\}$ for an undirected signed graph $G = (V, E, x)$ expresses each triangle as an *unordered* set of edges. However, we implicitly impose an ordering on the edges of t when we form $x_t = (x_{e_1}, x_{e_2}, x_{e_3})$ because the vector x_t is an ordered object. As this ordering is arbitrary, $d : \{0,1\}^3 \to \mathbb{R}_+$ must be invariant to the order of its arguments, i.e., for any permutation matrix π, $d(x_t) = d(x_t \pi)$. This requirement implies that there are only 4 possible values that d can take on: $d(0,0,0), d(1,0,0) = d(0,1,0) = d(0,0,1), d(0,1,1) = d(1,0,1) = d(1,1,0)$, and $d(1,1,1)$. We always use this restriction when learning or specifying d.

2 Extension of the model to directed graphs (Sec. 3.2)

We now describe the TRIANGLE BALANCE problem for a directed signed graph $G = (V, E, x)$. Since there is no global reference node, edge directionality does not influence the edge cost function $c(x_e, p_e) = \lambda_1(1 - p_e)x_e + \lambda_0 p_e(1 - x_e)$. However, when forming our triangle cost function, we now distinguish between two basic types of triangles: acyclic and cyclic. Since each triangle t is specified as an unordered set of edges, t is isomorphic to one of these basic types. We allow for different costs for each triangle type/sign combination by extending the definition of d to take the triangle type, $\sigma_t \in \{\text{CYCLIC}, \text{ACYCLIC}\}$, as an argument. Our overall objective is then

$$\sum_{e \in E} c(x_e, p_e) + \sum_{t \in T} d(x_t, \sigma_t). \tag{1}$$

Similarly to the undirected case, symmetries exist that d must take into account. When t is cyclical, the order of the edges does not matter and the same symmetries as in the undirected case must be obeyed. However, when t is acyclical each directed edge plays a different role in the triangle, and order matters. As such, we will assume that x_t is formed by always taking edges in the same canonical order when t is acyclical. Moreover, d takes on 8 distinct values when its second argument is ACYCLIC, yielding a total of 12 distinct values for d when G is directed.

3 NP-hardness of TRIANGLE BALANCE (Sec. 3.3)

We prove that TRIANGLE BALANCE is NP-hard for undirected graphs. It then follows immediately that the problem is NP-hard for directed graphs, too, since every undirected graph can be expressed as a directed graph,[1] which implies that the directed case is at least as hard as the undirected case.

[1] For instance, by fixing some order $v_1, \ldots, v_n$ on the vertices $V = \{v_1, \ldots, v_n\}$ and representing an undirected edge $e = \{v_i, v_j\}$ as the directed edge (v_i, v_j) if $i < j$, and (v_j, v_i) otherwise.

Further, in our problem formulation in Sec. 3.2 (Eq. 2) of the paper, the cost of an edge e, $c(x_e, p_e)$, is defined as a mere shortcut for $\lambda_1(1-p_e)x_e + \lambda_0 p_e(1-x_e)$, where $p_e \in [0,1]$, $\lambda_1 \in \mathbb{R}_+$, and $\lambda_0 \in \mathbb{R}_+$ are free parameters associated with the problem instance, and $x_e \in \{0,1\}$ is the optimization variable. That is, an edge e's cost of being positive (negative) is entirely determined by the sentiment model's output p_e and λ_1 (λ_0). For ease of exposition, we will first assume that edge costs are defined directly on a per-edge basis; i.e., each edge e has costs $c(x_e, e)$ associated immediately with it (for $x_e \in \{0,1\}$), without assuming parameters p_e, λ_1, and λ_0. We write $c_1(e)$ for $c(1,e)$, and $c_0(e)$ for $c(0,e)$. Having proved NP-hardness for this more general case, we will provide an argument why NP-hardness even holds for the more special case where edge costs are constrained to be of the form $\lambda_1(1-p_e)x_e + \lambda_0 p_e(1-x_e)$.

The TRIANGLE BALANCE problem is therefore defined as follows.

Problem. TRIANGLE BALANCE

Instance.

- An undirected graph $G = (V, E)$,
- a cost function $c_1 : E \to \mathbb{R}$, specifying the cost of each edge if it is positive,
- a cost function $c_0 : E \to \mathbb{R}$, specifying the cost of each edge if it is negative,
- a triangle cost function $d : \{0,1\}^3 \to \mathbb{R}$, where $d(x_t) = d(x_1, x_2, x_3)$ defines the cost of a triangle t whose edges have the signs $x_t = (x_1, x_2, x_3)$.

Task. Find edge signs $x \in \{0,1\}^{|E|}$ of minimum cost

$$H(x) = H_E(x) + H_T(x), \tag{2}$$

where the total cost $H(x)$ is the sum of the total edge cost

$$H_E(x) = \sum_{e \in E} x_e c_1(e) + (1-x_e) c_0(e) \tag{3}$$

and the total triangle cost

$$H_T(x) = \sum_{t \in T} d(x_t). \tag{4}$$

Theorem 1. TRIANGLE BALANCE *is NP-hard.*

Proof. We prove NP-hardness by reduction from TWO-LEVEL SPIN GLASS, which is known to be NP-hard and defined as follows [Bar82].

Problem. TWO-LEVEL SPIN GLASS

Instance.

- A two-level grid (cf. Fig. 1) $\bar{G} = (\bar{V}, \bar{E})$,
- an edge cost function $\bar{c} : \bar{E} \to \{-1, 0, +1\}$.

Task. Find vertex signs $\bar{x} \in \{-1,+1\}^{|V|}$ of minimum cost

$$\bar{H}(\bar{x}) = -\sum_{(u,v) \in \bar{E}} \bar{c}(u,v)\, \bar{x}_u \bar{x}_v. \tag{5}$$

Figure 1: A two-level grid [Bar82]. © Sol LeWitt: Cubic-Modular Wall Structure, Black, 1966, Museum of Modern Art, New York.

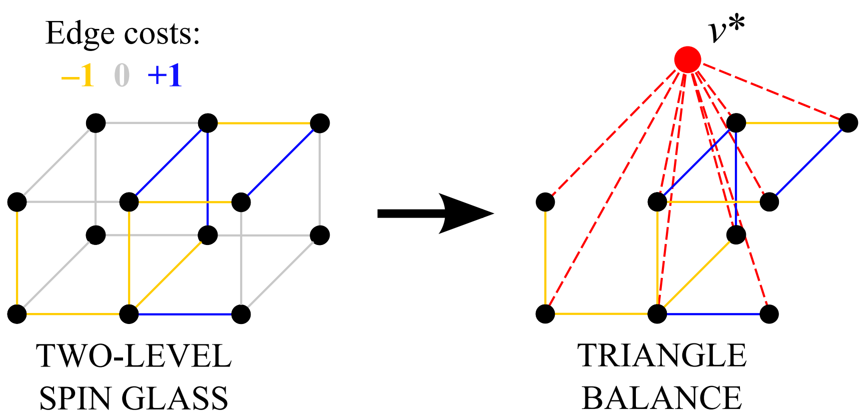

Figure 2: Illustration of the reduction from TWO-LEVEL SPIN GLASS to TRIANGLE BALANCE. (Dangling vertices with no incident blue or yellow edges not shown on right.)

As a matter of notation, for a graph $G = (V,E)$, we assume some order $v_1,\dots,v_n$ on the vertices $V = \{v_1,\dots,v_n\}$ and represent an undirected edge $e = \{v_i,v_j\}$ as the pair (v_i,v_j) if $i < j$, and (v_j,v_i) otherwise.

Here is a table of correspondences that motivates our reduction from TWO-LEVEL SPIN GLASS to TRIANGLE BALANCE:

TWO-LEVEL SPIN GLASS	TRIANGLE BALANCE
vertex labels	edge labels
vertex costs (all-zero)	edge costs
edge costs	triangle costs

So the idea is to map vertices from the TWO-LEVEL SPIN GLASS instance to edges in the TRIANGLE BALANCE instance, and edges in the TWO-LEVEL SPIN GLASS instance to triangles in the TRIANGLE BALANCE instance. The reduction is illustrated in Fig. 2. Formally, we transform a TWO-LEVEL SPIN GLASS instance $(\bar{G} = (\bar{V},\bar{E}),\bar{c})$ into a TRIANGLE BALANCE instance $(G = (V,E),c_1,c_0,d)$ as follows:

$$
\begin{aligned}
V &= \bar{V} \cup \{v^*\}, \quad \text{where } v^* \notin \bar{V} & (6)\\
E &= \{e \in \bar{E} : \bar{c}(e) \neq 0\} \cup \{(v, v^*) : v \in \bar{V}\} & (7)\\
c_1(e) &= \begin{cases} 0 & \text{if } e \notin \bar{E} \\ 0 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = +1 \\ 2|\bar{E}|+1 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = -1 \end{cases} & (8)\\
c_0(e) &= \begin{cases} 0 & \text{if } e \notin \bar{E} \\ 0 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = -1 \\ 2|\bar{E}|+1 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = +1 \end{cases} & (9)\\
d(x_1, x_2, x_3) &= \begin{cases} -1 & \text{if } x_1 + x_2 + x_3 \in \{1, 3\} \\ +1 & \text{if } x_1 + x_2 + x_3 \in \{0, 2\} \end{cases} & (10)
\end{aligned}
$$

We need to prove that each optimal solution to the original problem induces an optimal solution to the transformed problem, and *vice versa*. Our first observation is that, in the transformed problem, there is

1. **exactly one new edge** $(v, v^*) \in E \setminus \bar{E}$ **for each old vertex** $v \in \bar{V}$ (red edges in Fig. 2), creating
2. **exactly one new triangle** $\{(v_1, v_2), (v_1, v^*), (v_2, v^*)\}$ **for each old non-zero edge** $(v_1, v_2) \in E \cap \bar{E}$.

Further, since every two-level grid $\bar{G}$ is triangle-free, the set of triangles in G corresponds exactly to the set of non-zero edges in $\bar{G}$.

Due to item 1, there is a one-to-one correspondence between labelings of the old (black in Fig. 2) vertices $\bar{V}$ and labelings of the new (red in Fig. 2) edges $E \setminus \bar{E}$. Further, the old edge cost function $\bar{c}$ induces a labeling of the old (blue and yellow) edges $E \cap \bar{E}$. So, for each labeling $\bar{x}$ of the old vertices $\bar{V}$, we can define exactly one corresponding labeling x of the edges E of the transformed problem:

$$
x_{(u,v)} = \begin{cases} \bar{x}_u & \text{if } (u, v) \in E \setminus \bar{E}, \text{ i.e., if } v = v^*, \\ \bar{c}(u, v) & \text{if } (u, v) \in E \cap \bar{E}. \end{cases} \tag{11}
$$

We will show below that x has cost $H(x) = \bar{H}(\bar{x})$. Also, any x' that labels the old edges $E \cap \bar{E}$ differently than (11) will have cost $H(x') \geq 2|\bar{E}| + 1 - |\bar{E}| = |\bar{E}| + 1 > \bar{H}(\bar{x}) = H(x)$, due to (10) and the third case of (9), and because the cost $\bar{H}(\bar{x})$ in the original problem is never more than $|\bar{E}|$. Hence, any optimal solution x^* must label the old edges $E \cap \bar{E}$ as in (11), which means that each optimal solution $\bar{x}^*$ to the old problem is in a one-to-one correspondence with an optimal solution x^* to the transformed problem, and $\min_x H(x) = H(x^*) = \bar{H}(\bar{x}^*) = \min_{\bar{x}} \bar{H}(\bar{x})$.

To see that $H(x) = \bar{H}(\bar{x})$ for the $\bar{x}$-to-x transformation of (11), first observe that the total edge cost $H_E(x)$ (cf. (3)) is 0, since new edges are always free (case 1 of (9)) and old edges are free if they are signed according to $\bar{c}$ (case 2 of (9)).

Regarding the triangle cost $H_T(x)$ (cf. (4)), observe that a triangle (which always consists of one old edge and two new edges) has an even number of negative edges if and only if its old edge is positive and its two new edges are of the same sign, or its old edge is negative and its two new edges are of opposite signs. Therefore (cf. (10)), the number of triangles with cost ± 1 equals the number of edges (u, v) of the original problem for which $\bar{c}(u, v)\bar{x}_u\bar{x}_v = \mp 1$, or equivalently, $-\bar{c}(u, v)\bar{x}_u\bar{x}_v = \pm 1$. So the total triangle cost $H_T(x)$ of the transformed problem equals the total cost $\bar{H}(\bar{x})$ of the original problem (cf. (5)).

Adding edge and triangle costs, we obtain the total cost of the transformed problem as $H(x) = H_E(x) + H_T(x) = 0 + \bar{H}(\bar{x}) = \bar{H}(\bar{x})$.

To conclude, we show that the problem is NP-hard even when we constrain the edge costs $c(x_e, p_e)$ to be of the form $\lambda_1(1 - p_e)x_e + \lambda_0 p_e(1 - x_e)$, respectively, as assumed in Eq. 2 of the paper (cf. the discussion at the beginning of this section). In this case, instead of defining $c_1(e)$ and $c_0(e)$ for each $e \in E$, we need to define p_e for each $e \in E$, as well as λ_1 and λ_0. We define

$$\begin{aligned} p_e &= \begin{cases} \frac{1}{2} & \text{if } e \notin \bar{E} \\ 1 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = +1 \\ 0 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = -1 \end{cases} \\ \lambda_1 &= 2|\bar{E}|+1, \\ \lambda_0 &= 2|\bar{E}|+1. \end{aligned}$$

Since $c_1(e) = c(1, p_e) = \lambda_1(1-p_e)$ and $c_0(e) = c(0, p_e) = \lambda_0 p_e$, we can simply use the following edge costs in the same proof as above:

$$c_1(e) = \begin{cases} |\bar{E}|+\frac{1}{2} & \text{if } e \notin \bar{E} \\ 0 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = +1 \\ 2|\bar{E}|+1 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = -1 \end{cases} \tag{12}$$

$$c_0(e) = \begin{cases} |\bar{E}|+\frac{1}{2} & \text{if } e \notin \bar{E} \\ 2|\bar{E}|+1 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = +1 \\ 0 & \text{if } e \in \bar{E} \text{ and } \bar{c}(e) = -1 \end{cases} \tag{13}$$

The same proof as above will be valid, with the small difference that now the costs of the optimal solutions to the TWO-LEVEL SPIN GLASS and TRIANGLE BALANCE instances will differ by a constant $|\bar{V}|(|\bar{E}|+\frac{1}{2})$, as there are $|\bar{V}|$ edges in $E \setminus \bar{E}$, each of which contributes a cost of $|\bar{E}|+\frac{1}{2}$. $\square$

A final remark: Although TRIANGLE BALANCE was defined for general triangle-cost functions d in Eq. 1 of the paper, our reduction shows that even the special case of d that rewards triangles with an even, and punishes triangles with an odd, number of negative edges is NP-hard. This is noteworthy since it is the notion of social balance originally considered in sociology [Hei46].

References

[Bar82] Francisco Barahona. On the computational complexity of Ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241–3253, 1982.

[Hei46] Fritz Heider. Attitudes and cognitive organization. *The Journal of Psychology*, 21(1):107–112, 1946.

[WPLP14] Robert West, Hristo S. Paskov, Jure Leskovec, and Christopher Potts. Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics*, 2, 2014.