# Energy Models for Drawing Signed Graphs

Anne-Marie Kermarrec, Afshin Moin

# Energy Models for Drawing Signed Graphs*

Anne-Marie Kermarrec and Afshin Moin
INRIA Rennes Bretagne Atlantique, France

**Abstract**

Graph drawing is the pictorial representation of graphs in a multi-dimensional space. Energy models are the prevalent approach to graph drawing. In this paper, we propose energy models for drawing *signed* unidirectional graphs where edges are labeled either as positive (attractive) or as negative (repulsive). The existent energy models do not discriminate against edge sign. Hence, they do not lend themselves to drawing signed graphs. We suggest a *general equation for signed energy models* by proposing a *dual energy model* for graphs containing uniquely negative edges, and combining it linearly with the primary model. We then concentrate on revealing the community structure of social network graphs (sociograms) where edge sign represents the state of relationship between two individuals. In this goal, *Signed LinLog* model is built based on *LinLog* model whose clustering properties for unsigned graphs is already known. The properties of Signed LinLog model are outlined analytically, and its synthetic and real layouts are presented.

## 1 Introduction

The drawing of a graph called *layout* is the graphical representation of the graph where each vertex is assigned with a coordinate vector in a metric space. Graph drawing has applications in many branches of science needing information visualization like cartography, bioinformatics and social network analysis. Large scale Internet applications can also benefit from graph drawing to assign hosts with virtual network coordinates reflecting Internet latency [3].

*Energy-based* (also known as *force-based*[1]) technique [9, 5, 11, 6, 4, 19] is a pretty popular approach to graph drawing thanks to advantages like good layout quality, ease of implementation and code flexibility. The basic assumption in these models [5, 6, 4, 19] is that there is an *attraction* force between the connected vertices while all vertices repulse each other. This line of modeling captures all edges as a measure of attraction, while in many contexts like social networks [8], vertices (users) express positive or *negative* attitude towards each other. Consequently, the existing models place connected vertices close to each other even if some edges reflect dislike. While numerous works have studied positive interactions in social networks, there has been less research about signed networks [16]. In particular, there is a big gap in

*This work is supported by the ERC Starting Grant GOSSPLE number 204742.

[1]Force is the minus gradient of energy. Hence, the minimum energy in a system is equivalent to the force equilibrium.

the literature of *drawing* signed graphs. We are only aware of the work in [14] shortly investigating prediction, clustering and visualization in signed graphs by generalizing the spectral approaches for signed graphs. However, their visualization technique has not the clustering property we seek. This lack of literature about signed graph visualization motivated us to suggest *energy models for drawing unidirectional signed graphs.* In these models, positive edges are interpreted as attraction between vertices while negative edges denote repulsion. In this paper we first give the summary of the related work in 2. Notations are defined in Section 3. In Section 4, we suggest a dual formulation for a popular class of energy models aimed at drawing graphs containing only negative edges. This model is linearly combined with its primary to suggest a *general equation* for signed energy functions. Precisely because we are interested in drawing sociograms where clusters indicate communities, we emphasize how to draw *clustered* layouts of signed graphs. The clustering property of energy models has recently been elaborated by Noack [19]. In sprite of this work, we suggest the *Signed LinLog* model, and analytically derive its clustering properties. The behavior of the model is clarified in Section 5 presenting a toy example as well as layouts of real data. Section 9 concludes the paper.

## 2 Related Work

The problem of drawing unsigned graphs has been addressed in several works [9, 5, 11, 6, 4, 19]. Energy-based technique is the prevalent approach to graph drawing. This approach has several advantages like good layout quality, ease of implementation and flexibility (easily modifiable codes). It is composed of two components: the energy model and the optimization algorithm. The energy model is a heuristic function defining the energy of layouts. The optimal layout is that with minimum energy. The role of the optimization algorithm is to find the *local* minimum of the energy function. Energy-based approaches differ in the energy function and the optimization algorithm they use. Two main disadvantages associated with energy models are the difficulty of finding the *global* minimum of the energy function and high running time for very large graphs. In this approach, vertex coordinates are determined by minimizing a heuristic energy function through an iterative optimization algorithm. The Spring Embedder method [5] was one of the earliest variants of energy models in which a graph is considered as a collection of electrically charged rings connected through spring edges. Therefore, all vertices repulse each other, while the attraction exists uniquely between connected vertices. The equilibrium is sought between all attractive and repulsive forces by iteratively modifying vertex coordinates, minimizing the energy after a number of cycles. The energy minimization algorithm of this work had high complexity rendering it inefficient for large graphs. Fruchterman and Reingold [6] refined this algorithm by first computing the net force exerted on each vertex, then moving it in the direction of the net force by an extent limited by a decreasing function of cycle number. The scalability was further enhanced by Tunkelang [22]. The author adopts the algorithm by Barnes and Hut [1] using a quad-tree to approximate the repulsion force of many distant vertices as a single one. In this paper, we also use this algorithm to optimize our energy functions. Kamada and Kawai [11] model the graph as a system of springs acting according to Hook's Law. A spring is assumed between every two vertices whose rest length is proportional to the graph-theoretical distance between the end-points and whose spring constant is inversely proportional to the square of this distance. Hall [9, 12] introduces a quadratic energy function for connected vertices, but puts geometrical constraints on the vertex coordinates in order to prevent them from overlapping on a single point. Lately, Koren and Çivril [13] have proposed a binary stress model by linking up the stress and the electrical spring models. A survey of force-based models is available in [2].

All these works have the goal of generating layouts in a way that the readability criteria for aesthetic drawing is satisfied. These criteria are in general uniform distribution of vertices, minimum edge crossing and uniform edge length. They lead to increased layout clarity, but prevent the appearance of clusters. Recently, Noack [19] has elaborated the problem of clustering energy models. The author argues that energy models are in general not clustering, and suggests the *LinLog* model as the special model drawing layouts where convex subgraphs are plot separately, and the distance between clusters of a bipartition decreases with the number of connecting edges.

None of the works described above refer to signed graph drawing. To the best of authors' knowledge, there is only [14] addressing shortly the problem of signed graph drawing using spectral approaches. Though, their approach does not have our desired clustering property.

## 3 Definitions

A d-dimensional layout $p$ of a graph $G = (V, E)$ is a mapping of vertices into a Euclidean space $M : V \longrightarrow R^d$ where $\forall v \in V$ is assigned with a coordinate vector $p_v$. The Euclidean distance between $u$ and $v$ is denoted by $\|p_u - p_v\|$. Here, we clarify the clustering property by introducing the notions of *density* and *distance*. We denote unsigned graphs by $G = (V, E)$ where $V$ is the set of vertices and $E$ the set of edges. A subgraph of an unsigned graph is considered as a cluster, if it contains lots of internal edges while few edges to the rest of the graph. There exist many different clusterings for a particular graph. To judge the clustering quality between two disjoint clusters, a *coupling measure* is defined between them. The best clustering is the one minimizing the coupling. One widely used coupling measure between two clusters $V_1$ and $V_2$ is the *cut*:

$$cut[V_1, V_2] = \left| E_{V_1 \times V_2} \right|,$$

where $E_{V_1 \times V_2}$ represents the set of edges between $V_1$ and $V_2$. The cut has the disadvantage of preferring biased clusters, i.e. one huge cluster against a tiny one. The *normalized cut* also known as *density* can remove this drawback:

$$density(V_1, V_2) = \frac{cut[V_1, V_2]}{|V_1|\,|V_2|}.$$

Cut and density are coupling measures for generation of graph bipartitions[2]. Other measures like *Newman Modularity* [17] may be used to generate clusterings with more than two clusters. We present three definitions of distance for $F \subset V^{(2)}$ and layout $p$ widely used in graph

[2] $(B_1, B_2)$ is called a *bipartition* of $G(V, E)$, if $B_1 \cap B_2 = \emptyset$ and $B_1 \cup B_2 = V$.

geometry. The *arithmetic mean distance* is the linear average of the Euclidean distance between all vertex pairs:

$$arith(F,p) = \frac{1}{|F|} \sum_{\{u,v\}\in F} \|p_u - p_v\|.$$

The *geometrical mean distance* and the *harmonic mean distance* of $F$ are non-linear averaging functions to compute distance:

$$geo(F,p) = \sqrt[|F|]{\prod_{\{u,v\}\in F} \|p_u - p_v\|}, \qquad harm(F,p) = \frac{|F|}{\sum_{\{u,v\}\in F} \frac{1}{\|p_u - p_v\|}}.$$

In clustered layouts the vertices of each cluster are plot close to each other. For unsigned graphs with *attractive* edges the distance between clusters must be in *inverse* relation with their density. For unsigned graphs with *repulsive* edges, where a cluster contains vertices having negative edges towards another subset of vertices, the distance between clusters must vary directly with their density.

Here, we generalize the above definitions to signed graphs. We represent signed graphs with $G_s = (V,E)$, where $E = E^+ \cup E^-$, $E^+ \cap E^- = \emptyset$. The set of positive and negative edges are denoted by $E^+$ and $E^-$, respectively. In signed graphs, a cluster is a subgraph containing a large number of internal positive edges and few internal negative edges. An optimal bipartition of a signed graph maximizes both the number of positive internal edges of the partitions and the number of negative edges between them. The positive and negative cut between two disjoint clusters $V_1,\ V_2 \subset V$ indicate the number of connecting positive and negative edges, respectively:

$$cut^+[V_1,V_2] = \left|E^+_{V_1\times V_2}\right|, \quad cut^-[V_1,V_2] = \left|E^-_{V_1\times V_2}\right|.$$

Consequently, *positive density* and *negative density* between two disjoint clusters are defined as:

$$density^+(V_1,V_2) = \frac{cut^+[V_1,V_2]}{|V_1|\,|V_2|}, \qquad density^-(V_1,V_2) = \frac{cut^-[V_1,V_2]}{|V_1|\,|V_2|}.$$

In this paper, we also introduce three new definitions of distance for signed layouts appearing further in the theorems. The *positive arithmetic mean distance* for $F \subset V^{(2)}$ and layout $p$ is defined as:

$$\begin{aligned} arith^+(F,p) &= \frac{\sum_{\{u,v\}\in F} \lambda_{uv}\|p_u - p_v\|}{\left|E^+_F\right| + |F|}, \quad \lambda_{uv} = \begin{cases} 2 & if\ \{u,v\} \in E^+_F \\ 1 & otherwise \end{cases} \\ &= \frac{\left|E^+_F\right| arith(E^+_F,p) + |F|\, arith(F,p)}{\left|E^+_F\right| + |F|}, \end{aligned}$$

where $E^+_F$ is the set of positive edges of $F$. It is a linear averaging function where positively-connected vertex pairs are counted twice. We define in the same way the *negative geometrical* *mean distance*:

$$\begin{aligned} geo^-(F,p) &= \sqrt[\left|E_F^-\right|+|F|]{\prod_{\{u,v\}\in F} \lambda'_{uv}\|p_u - p_v\|}, \quad \lambda'_{uv} = \begin{cases} 2 & if\ \{u,v\} \in E_F^- \\ 1 & otherwise \end{cases} \\ &= \sqrt[\left|E_F^-\right|+|F|]{geo(E_F^-,p)^{\left|E_F^-\right|} geo(F,p)^{|F|}}, \end{aligned}$$

where $E_F^-$ is the set of negative edges of $F$. Finally, the *negative harmonic mean distance* of $F$ on layout $p$ is defined as:

$$\begin{aligned} harm^-(F,p) &= \frac{\left|E_F^-\right| + |F|}{\sum_{\{u,v\}\in F} \frac{\lambda'_{uv}}{\|p_u - p_v\|}}, \quad \lambda'_{uv} = \begin{cases} 2 & if\ \{u,v\} \in E_F^- \\ 1 & otherwise \end{cases} \\ &= \frac{\left|E_F^-\right| + |F|}{\frac{\left|E_F^-\right|}{harm(E_F^-)} + \frac{|F|}{harm(F,p)}}. \end{aligned}$$

Negative edges between $V_1$ and $V_2$ are weighted twice as much in $geo^-(F,p)$ and $harm^-(F,p)$. These definitions can be generalized further (as in Corollary 4.1 and Corollary B.2) to count positive and negative edges or disconnected pairs as many times as desired. In clustered signed layouts, the distance between two clusters increases with negative density and decreases with positive density.

## 4 Energy Models for Clustering Signed Graphs

The conventional work line in the designation of energy models is to suppose disconnected vertices repulse each other until infinity. The attraction force exists wherever an edge exists between vertices, and the system rests when the equilibrium between these forces is achieved. For layout $p$ of an unsigned graph $G = (V, E^+)$ with *attractive* edges, many of the known energy models [19, 6, 4] have the following form:

$$U = \sum_{\{u,v\}\in E^+} f(\|p_u - p_v\|) + \sum_{\{u,v\}\in V^{(2)}} g(\|p_u - p_v\|),$$

where $f(\|p_u - p_v\|)$ represents the attraction energy, and $g(\|p_u - p_v\|)$ the repulsion energy. In this conventional work line, disconnected subgraphs repulse each other towards infinity. Reminding the small world phenomenon in human societies [15], this infinite repulsion is questionable in social network visualization, as there is no convincing reason to count two disconnected communities of people as infinitely far from each other. In other words, the absence of interactions between two communities suggests ignorance rather than antagonism.

We introduce a new category of energy models which can be viewed as the dual formulation of the actual models. The idea is that each vertex establishes links to other vertices only based on a *distrust* relation. In contrast to the conventional models, we assume there is an attraction force between every two different vertices while connected vertices repulse each other. For layout $p$ of an unsigned graph $G = (V, E^-)$ with *repulsive* edges, this dual model has the general form of:

$$U_{Dual} = \sum_{\{u,v\} \in V^{(2)}} f(\|p_u - p_v\|) + \sum_{\{u,v\} \in E^-} g(\|p_u - p_v\|).$$

The particular advantage of this model is to separate clusters of vertices based on their similar dislike habits. In a connected bipartite graph for example, the two partitions of vertices are plot apart from each other, with the vertices of each partition overlapping on the same point. Though, this model may be criticized in the inverse way of conventional models. Namely, there is no reason for two disconnected vertices to be extremely close to, even overlapping, each other. In addition, drawing models are required to avoid vertex overlap as one of the primitive requirements in graph embedding. Fortunately, a linear combination of conventional models with their dual can lift the drawbacks of both, and satisfy the vertex non-overlap condition. In addition, while primary and dual models can handle either positive or negative edges, the combined version upgrades the modeling power to signed graphs. For a signed graph $G_s = (V, E)$ we define:

$$\begin{aligned} U_{SignedGraph} \quad = & \sum_{\{u,v\} \in E^+} f(\|p_u - p_v\|) + \sum_{\{u,v\} \in E^-} g(\|p_u - p_v\|) + \\ & \sum_{\{u,v\} \in V^{(2)}} f(\|p_u - p_v\|) + g(\|p_u - p_v\|). \end{aligned} \tag{1}$$

Other possible forms are suggested in Section 6. To have an intuition into the behavior of the model, lets consider the simple example of a graph with two vertices. If the vertices are disconnected, the minimum energy of the system is a solution to $f(\|p_u - p_v\|) + g(\|p_u - p_v\|) = 0$. In case they are positively or negatively connected, the minimum energy layout would correspond to $2f(\|p_u - p_v\|) + g(\|p_u - p_v\|) = 0$ and $f(\|p_u - p_v\|) + 2g(\|p_u - p_v\|) = 0$, respectively. It is seen from this simple example that the model places disconnected vertices within a neutral distance from each other, while positively/negatively connected vertices are put closer/further w.r.t. the neutral distance. The ability to draw disconnected graphs is an advantage of signed energy models as conventional energy models have difficulties with handling disconnected graphs. Recalling we are interested in clustered layouts, we focus on the Signed LinLog model whose clustering properties of the unsigned version is proven in [19]. LinLog [19] and Dual LinLog models are defined as follows:

$$U_{LinLog} = \sum_{\{u,v\} \in E^+} \|p_u - p_v\| - \sum_{\{u,v\} \in V^{(2)}} \ln \|p_u - p_v\|.$$

$$U_{DualLinLog} = \sum_{\{u,v\} \in V^{(2)}} \|p_u - p_v\| - \sum_{\{u,v\} \in E^-} \ln \|p_u - p_v\|.$$

Signed LinLog is then defined as:

$$U_{SignedLinLog} = \sum_{\{u,v\}\in E^+} \|p_u - p_v\| - \sum_{\{u,v\}\in E^-} \ln \|p_u - p_v\| + \sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| - \ln \|p_u - p_v\|. \tag{2}$$

In the rest of this section we derive the clustering properties of this new model. Theorem 4.1 states that a layout with minimum Signed LinLog energy is the best trade-off between the minimization of the positive arithmetic mean and the maximization of the negative geometrical mean of the whole vertices. The former necessitates putting the vertices close to each other, while the latter is likely to move them away. The point is that positive edges have more weight in $arith^+(V^{(2)}, p)$, but negative edges are weighted more in $geo^-(V^{(2)}, p)$. Then, shortening positive edges results in more decrease in $arith^+(V^{(2)}, p)$ while lengthening negative edges causes more increase in $geo^-(V^{(2)}, p)$. Consequently, positive edges become shorter in average than the mean distance between all vertices, and negative edges become longer than that. This results in friends lying close to each other while foes move apart.

**Theorem 4.1** *If $p^0$ is a drawing of the signed graph $G_s = (V, E)$ with minimum Signed LinLog energy, $p^0$ also minimizes $\frac{arith^+(V^{(2)},p)}{geo^-(V^{(2)},p)}$.*

**Proof** Let $p^0$ be a layout with minimum Signed LinLog energy.
If $\sum_{\{u,v\}\in E^+} \|p^0_u - p^0_v\| + \sum_{\{u,v\}\in V^{(2)}} \|p^0_u - p^0_v\| = c$, then $p^0$ is a solution to:

$$\begin{aligned} &\text{minimize}\Big(-\sum_{\{u,v\}\in E^-} \ln\|p_u - p_v\| - \sum_{\{u,v\}\in V^{(2)}} \ln\|p_u - p_v\|\Big) \\ &\text{subject to} \quad \sum_{\{u,v\}\in E^+} \|p_u - p_v\| + \sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| = c. \end{aligned}$$

The above expression may be reformulated in the form of
$\text{minimize} - \ln(geo^{\left|E^-\right|}(E^-, p)geo^{\left|V^{(2)}\right|}(V^{(2)}, p))$.
Since $\sqrt[\left|E^-\right|+\left|V^{(2)}\right|]{\exp(x)}$ is an increasing function of $x$, the minimization of this expression is equivalent to $\text{minimize}\exp(\ln \frac{1}{geo^-(V^{(2)},p)})$. Multiply the numerator by the constant $arith^+(V^{(2)}, p)$, and rewriting the restriction, we obtain:

$$\text{minimize}\frac{arith^+(V^{(2)}, p)}{geo^-(V^{(2)}, p)} \quad \text{subject to} \quad arith^+(V^{(2)}, p) = \frac{c}{\left|E^+\right| + \left|V^{(2)}\right|}.$$

Suppose there exists a layout $q^0$ of $G_s$ with minimum Signed LinLog energy for which $\frac{arith^+(V^{(2)},q^0)}{geo^-(V^{(2)},q^0)} < \frac{arith^+(V^{(2)},p^0)}{geo^-(V^{(2)},p^0)}$.
We can always define a scaling $q^1 = \frac{c}{(\left|E^+\right|+\left|V^{(2)}\right|)arith^+(V^{(2)},q^0)} q^0$ for which $arith^+(V^{(2)}, q^1) =$

$\frac{c}{|E^+|+\left|V^{(2)}\right|}$, but $\frac{arith^+(V^{(2)},q^1)}{geo^-(V^{(2)},q^1)} = \frac{arith^+(V^{(2)},q^0)}{geo^-(V^{(2)},q^0)} < \frac{arith^+(V^{(2)},p^0)}{geo^-(V^{(2)},p^0)}$. This is a contradiction. Hence $q^0$ does not exist and the restriction may always be removed. ∎

Similar theorems exist for LinLog and Dual LinLog models:

**Theorem 4.2** *Let $G = (V, E^+)$ be a connected graph, and let $p^0$ be a layout of $G$ with minimum LinLog energy. Then $p^0$ also minimizes $\frac{arith(E^+,p)}{geo(V^{(2)},p)}$.*

**Theorem 4.3** *If $p^0$ is a drawing of a graph $G = (V, E^-)$ with minimum Dual LinLog energy, $p^0$ also minimizes $\frac{arith(V^{(2)},p)}{geo(E^-,p)}$.*

Adding multiplicative constants to the LinLog model has only a zooming effect, but does not change its minimum [18]. However, the minimum of Signed LinLog model depends on the constants. It can be proved in the same way as in Theorem 4.1 that:

**Corollary 4.1** *The minimization of Weighted Signed LinLog energy defined as*

$$\begin{aligned} U_{WeigthedSignedLinLog} &= \sum_{\{u,v\}\in E^+} k_1\|p_u - p_v\| - \sum_{\{u,v\}\in E^-} k_3 \ln \|p_u - p_v\| + \\ &\quad \sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| - k_3 \ln \|p_u - p_v\|, \end{aligned}$$

*is equivalent to the minimization of $\frac{arith^{k_1+}(V^{(2)},p)}{geo^{k_2-,k_3d}(V^{(2)},p)}$.*

$arith^{k_1+}(V^{(2)}, p)$ is defined similar to $arith^+(V^{(2)}, p)$, but positive edges are counted $(k_1 + 1)$ times. In the same way, disconnected pairs are weighted $k_3$ times as much in $geo^{k_2-,k_3d}(V^{(2)}, p)$, while negative edges are counted $(k_2 + k_3)$ times. These three constants determine *how much* positive/negative edges are shorter/longer in average w.r.t. the mean neutral distance between all vertices. Specifically in large sparse graphs, encountered frequently in social networks, the adjustment of these parameters is helpful to improve the revelation of clusters.

While Theorem 4.1 explains why Signed LinLog reveals the clusters, Theorem 4.4 is about the distance interpretability in bipartition layouts. It states that the negative harmonic mean distance between two partitions of a signed graph in a one-dimensional layout with minimum Signed LinLog energy varies directly with the negative density between them, and inversely with their positive density.

**Theorem 4.4** *Let $p^0$ be a one-dimensional drawing of the signed graph $G_s = (V, E)$. Let $(V_1, V_2)$ be a bipartition of $V$ such that the vertices in $V_1$ have smaller positions than the vertices in $V_2$ (i.e. $\forall v_1 \in V_1,\ \forall v_2 \in V_2 : p_{v_1} < p_{v_2}$). Then:*

$$harm^-(V_1 \times V_2, p^0) = \frac{1 + density^-(V_1, V_2)}{1 + density^+(V_1, V_2)}.$$

| Model | Minimization equivalence | One-dimensional bipartition |
|---|---|---|
| LinLog [19] | minimize $\frac{arith(E^+,p)}{geo(V^{(2)},p)}$ | $harm(V_1 \times V_2, p^0) = \frac{1}{density^+(V_1,V_2)}$ |
| Dual LinLog | minimize $\frac{arith(V^{(2)},p)}{geo(E^-,p)}$ | $harm(E^-_{V_1 \times V_2}, p^0) = density^-(V_1, V_2)$ |
| Signed LinLog | minimize $\frac{arith^+(V^{(2)},p)}{geo^-(V^{(2)},p)}$ | $harm^-(V_1 \times V_2, p^0) = \frac{1+density^-(V_1,V_2)}{1+density^+(V_1,V_2)}$ |

Table 1: Summary of the clustering properties of primary, dual and signed LinLog.

**Proof** Let $p^0$ be a layout with minimum Signed LinLog energy. If we add $d \in R$ to the coordinates of the vertices of $V_1$ in a way that the largest coordinate of the vertices in $V_1$ remain less than the smallest coordinate of the vertices in $V_2$, the Signed LinLog energy becomes:

$$\begin{aligned} U_{SignedLinLog}(d, p^0) &= \sum_{\{u,v\} \in E^+_{V_1^{(2)}} \cup E^+_{V_2^{(2)}}} |p_u - p_v| - \sum_{\{u,v\} \in \cup E^-_{V_1^{(2)}} \cup E^-_{V_2^{(2)}}} \ln |p_u - p_v| \\ &+ \sum_{\{u,v\} \in V_1^{(2)} \cup V_2^{(2)}} |p_u - p_v| - \ln |p_u - p_v| \\ &+ \sum_{\{u,v\} \in E^+_{V_1 \times V_2}} (|p_u - p_v| + d) - \sum_{\{u,v\} \in E^-_{V_1 \times V_2}} \ln(|p_u - p_v| + d) \\ &+ \sum_{\{u,v\} \in V_1 \times V_2} |p_u - p_v| + d - \ln(|p_u - p_v| + d). \end{aligned}$$

Since $p^0$ is a layout with minimum energy, the above function has a minimum at $d = 0$, i.e. $U'_{SignedLinLog}(d = 0, p^0) = 0$. Then:

$$\left| E^+_{V_1 \times V_2} \right| + |V_1 \times V_2| = \sum_{\{u,v\} \in E^-_{V_1 \times V_2}} \frac{1}{|p_u - p_v|} + \sum_{\{u,v\} \in V_1 \times V_2} \frac{1}{|p_u - p_v|}.$$

Replacing the right side with $\frac{|V_1||V_2| + \left| E^-_{V_1 \times V_2} \right|}{harm^-(V_1 \times V_2, p^0)}$ and $|V_1 \times V_2|$ with $|V_1|\,|V_2|$, the result is obtained. ∎

Although Theorem 4.4 does not generalize to more than one dimensions, it remains *approximately* true for 1+ dimensional layouts of *clusterizable bipartitions*. Refer to Appendix B for more details of the approximation. Notice for graphs containing a higher number of clusters, there is in general no 2D or 3D drawing satisfying the clustering property between *every* two clusters. Namely, there exists no configuration where the distance between every two clusters follow the clustering criterion, without conflicting the geometrical constraints imposed by the triangle inequality w.r.t. a third cluster. Equivalent theorems for LinLog and Dual LinLog models are:

**Theorem 4.5** *Let $G = (V, E^+)$ be a connected graph, and let $p^0$ be a one-dimensional drawing of $G$ with minimum LinLog energy. Let $(V_1, V_2)$ be a partition of $V$ such that the vertices in $V_1$ have smaller positions than the vertices in $V_2$ (i.e. $\forall v_1 \in V_1,\ \forall v_2 \in V_2 : p_{v_1} < p_{v_2}$). Then, $harm(V_1 \times V_2, p) = \frac{1}{density(V_1,V_2)}$.*

**Theorem 4.6** *Let $p^0$ be a one-dimensional drawing of a graph $G = (V, E^-)$ with minimum Dual LinLog energy. Let $(V_1, V_2)$ be a partition of $V$ such that the vertices in $V_1$ have smaller positions than the vertices in $V_2$ (i.e. $\forall v_1 \in V_1,\ \forall v_2 \in V_2 : p_{v_1} < p_{v_2}$). Then, $harm(E^-_{V_1 \times V_2}, p^0) = density(V_1, V_2)$.*

Table 1 summarizes the clustering properties of Signed LinLog with its primary and dual. It is seen that Signed LinLog properties is a trade-off between the properties of its building blocks. Proofs for Dual LinLog theorems are found in Appendix A. This appendix also provides theorems about the relationship between the length of edges with the number of vertices in LinLog, Dual LinLog and Signed LinLog models. Refer to [19] for the proofs of LinLog theorems.

# 5 Signed LinLog in Action

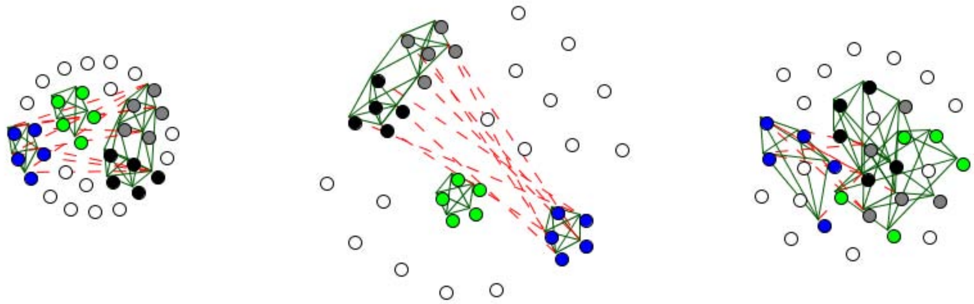
In this section we present a synthetic layout of Signed LinLog model as well as layouts of real data traces from Epinions [21], Slashdot [21], MovieLens [7] and tribal groups of Eastern Central Highlands of New Guinea [20].

## 5.1 Toy Examples

Figure 1b shows the Signed LinLog layout of a graph containing 4 clusters of 5 vertices each. The link probability inside clusters is 0.99 representing groups of close friends. The positive link probability between cluster 1 and 2 is 0.2, i.e. they have some common interests. The negative link probability between clusters 1 and 2 with 3 is 0.2, i.e. they distrust cluster 3 to some extent. Cluster 4 has no relationship with the other clusters. There are also 15 disconnected vertices. All clusters are clearly identifiable. Clusters 1 and 2 lie pretty close to each other, but far from cluster 3. Cluster 4 is found in a place with almost equal distance from the other clusters. Notice disconnected vertices take position in a way that their distance from other vertices remain as uniform as possible. We also implemented the signed versions of Fruchterman and Reingold [6] and Davidson and Harel [4] through Equation (1). Exact

| | | |
|---|---|---|
| Signed Fruchterman and Reingold | U = | $\sum_{\{u,v\}\in E^+} \frac{k_1}{3} \lvert p_u - p_v\rvert^3 - \sum_{\{u,v\}\in E^-} k_2 \ln \lvert p_u - p_v\rvert$ |
| | + | $\sum_{\{u,v\}\in V^{(2)}} (\frac{1}{3} \lvert p_u - p_v\rvert^3 - k_3 \ln \lvert p_u - p_v\rvert)$ |
| Signed Davidson and Harel | U = | $\sum_{\{u,v\}\in E^+} k_1 \lvert p_u - p_v\rvert^2 + \sum_{\{u,v\}\in E^-} \frac{k_2}{\lvert p_u - p_v\rvert^2}$ |
| | + | $\sum_{\{u,v\}\in V^{(2)}} (\lvert p_u - p_v\rvert^2 + \frac{k_3}{\lvert p_u - p_v\rvert^2}$ |
| Signed LinLog | U = | $\sum_{\{u,v\}\in E^+} k_1 \lvert p_u - p_v\rvert - \sum_{\{u,v\}\in E^-} k_2 \ln \lvert p_u - p_v\rvert$ |
| | + | $\sum_{\{u,v\}\in V^{(2)}} \lvert p_u - p_v\rvert - k_3 \ln \lvert p_u - p_v\rvert$ |

Table 2: Signed Energy Models

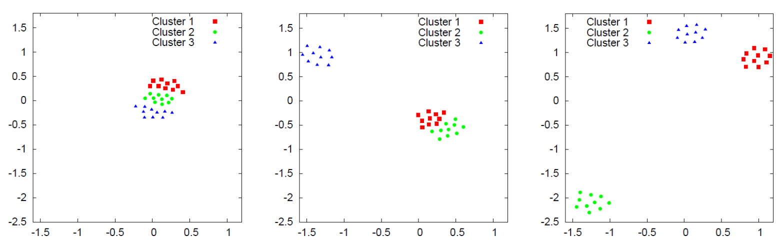


(a) Signed Fruchterman (b) Signed LinLog (c) Signed Davidson

Figure 1: Signed layouts of the toy example, cluster 1 is black, cluster 2 gray, cluster 3 blue and cluster 4 green (for all three models $k_1 = 3$, $k_2 = 2$, $k_3 = 2$). Positive edges are plot in green, and negative edges by dashed red lines. Signed LinLog has clear superiority in the representation of clusters.

equations are given in Table 2. As is obvious from Figures 1a and 1c, there is no way to detect the clusters once the vertices are discolored.

Here, we investigate the conformance of Signed LinLog layouts with the structural balance theory [10]. This theory originated in social psychology by F. Heider. It posits that triangles with three positive edges, or two negative edges and one positive edge are more likely to exist between three individuals. Hence, they should be more frequent in real human communities. In other words, "The friend of my enemy is my enemy" and "The friend of my friend is my friend" are more plausible than "The friend of my enemy is my friend" and "the enemy of my enemy is my enemy".

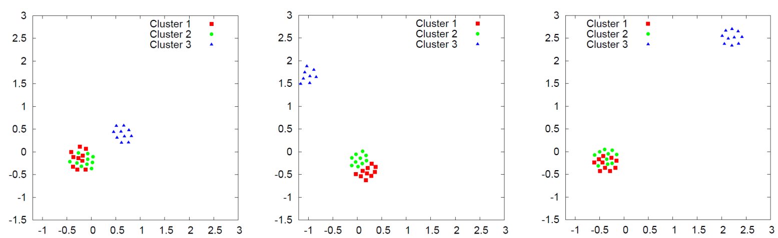



(a) $prob^+_{12} = 0.9$ and $prob^+_{23} = 0.9$ (b) $prob^+_{12} = 0.9$ and $prob^-_{23} = 0.9$ (c) $prob^-_{12} = 0.9$ and $prob^-_{23} = 0.9$

Figure 2: Conformance of Signed LinLog with Balance theory, $k_1 = k_2 = 3$, $k_3 = 1$.

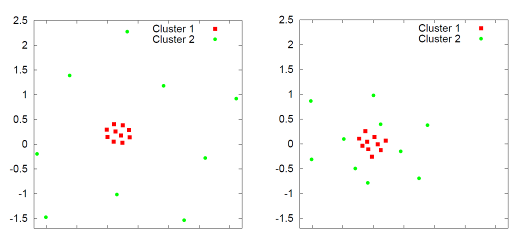
**Signed LinLog versus Balance Theory** Let a graph with three communities (clusters) of 10 friends ($prob^+_{ii} = 0.99$, $prob^-_{ii} = 0.0$). We triple the attractive and repulsive forces for better layout clarity ($k_1 = k_2 = 3$, $k_3 = 1$). In Figure 2a, cluster 1 and cluster 2, and cluster 2 and cluster 3 trust each other ($prob^+_{12} = 0.9$, $prob^+_{23} = 0.9$). The structural balance theory suggests that cluster 2 and cluster 3 trust each other with high probability. We can see that they are put close to each other suggesting the same idea. In Figure 2b, cluster 1, trusts cluster 2, but cluster 2 distrusts cluster 3. The balance theory predicts distrust between cluster 1 and cluster 3. It is seen that cluster 1 and 3 are far from each other in the layout. In Figure 2c cluster 1 distrusts cluster 2, and cluster 2 distrusts cluster 3. Balance theory predicts trust between cluster 1 and cluster 3. In the Signed LinLog layout however, cluster 3 and cluster 1 are not very close. They seem to keep a neutral position towards each. In other words, "the enemy of my enemy is *neither necessarily* my friend *nor* my enemy".

**My Enemy takes Her Friends Further from Me** Let a graph with three communities of 10 friends ($prob^+_{ii} = 0.99$, $prob^-_{ii} = 0.0$). This graph models three communities of friends. In all three figures there is strong trust between cluster 1 and cluster 2 explaining why they are almost merged. In Figure 3a the two clusters express no attitude towards cluster 3 ($prob^+_{12} = 0.9$, $prob^-_{13} = 0.0$, $prob^-_{23} = 0.0$). Both cluster 1 and cluster 2 lie in a neutral position with the same distance from cluster 3. In Figure 3b cluster 2 expresses distrust towards cluster 3, while there is no link between cluster 1 and cluster 3 ($prob^+_{12} = 0.9$, $prob^-_{13} = 0.0$, $prob^-_{23} = 0.9$). It is seen that both cluster 1 and cluster 2 lie further from cluster 3 in Figure 3b than in Figure 3a, but cluster 1 is still closer to cluster 3. In other words, cluster 1 moves away from cluster 3 although there is no explicit edge between them. This occurs because cluster 1 follows its friend, and this latter distrusts cluster 3. In Figure 3c where both cluster 1 and cluster 2 explicitly express their distrust towards cluster 3 ($prob^+_{12} = 0.9$, $prob^-_{13} = 0.9$, $prob^-_{23} = 0.9$), their distance gets more than the previous two cases. Both cluster 1 and cluster 2 have the same position against cluster 3.




(a) $prob^+_{12} = 0.9$ (b) $prob^+_{12} = 0.9$ and $prob^-_{23} = 0.9$ (c) $prob^+_{12} = 0.9$ and $prob^-_{13} = 0.9$ and $prob^-_{23} = 0.9$

Figure 3: Behavior of a cluster against the enemy of its friend, $k_1 = k_2 = 3$ and $k_3 = 1$




(a) $prob^+_{12} = 0.0$ (b) $prob^+_{12} = 0.6$

Figure 4: Behavior of a group of enemies, $k_1 = k_2 = 3$ and $k_3 = 1$

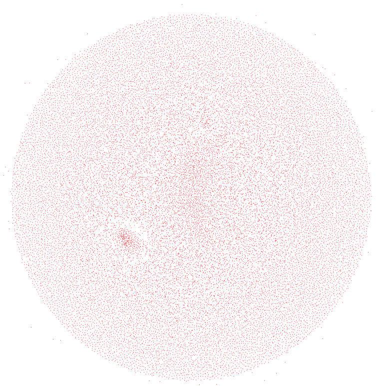
**Cluster of Enemies** This example considers the special case of a graph with one group of 10 enemies ($prob^+_{11} = 0.0, prob^-_{11} = 0.99$) and a cluster of 10 friends ($prob^+_{22} = 0.99, prob^-_{22} = 0.0$). Figure 4a shows the case where there is no link between two clusters ($prob^+_{12} = 0.0$, $prob^-_{12} = 0.0$). The enemies move away from each other but try to stay in a neutral position against cluster 1 by encircling it. In Figure 4b the enemies have in common that they trust cluster 1 ($prob^+_{12} = 0.6$). Then, they get closer to cluster 1 than in Figure 4a, but still try to be as far as possible from each other. It is interesting to note that the users of cluster 1 try to stay away from each other, but at the same time get as close as possible to cluster 2.

## 5.2 Real World layouts

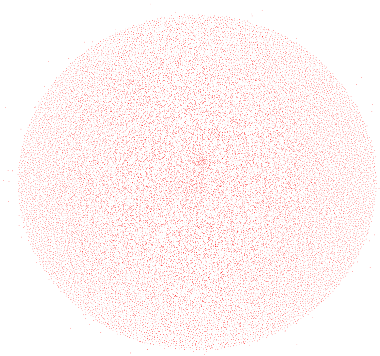
Epinions is a consumer review website where users can express trust or distrust in each other. The dataset consists of 841372 votes by 131828 voters. Slashdot [21] is a technology-related news website where a variety of user-generated information about latest technology and science-related news is circulated. Each user can mark others as friend if she likes their comments or as foe otherwise. The dataset contains 549202 votes by 82140 voters taken on the 21th of February 2009. Our statistical studies show that in the abolute majority of the cases, two users have both given the same vote for each other (either both positive or both negative). Despite, This conflict happens in 1% of two-way votes in Epinions dataset and 2% of them in Slashdot dataset. The layouts of Epinions and Slashdot are predented in Figures 5a and 5b, respectively. In the simulations, we first omitted the conflicting votes, then kept vertices with at least 5 positive incident unidirectional edges.

The tribal groups of Eastern Central Highlands of New Guinea is consisted of 16 tribes being in war or peace state with each other. Each cluster has friend or foe relationship with others, but there is no foe state inside the same cluster. The corresponding layout is given is Figure 5c.

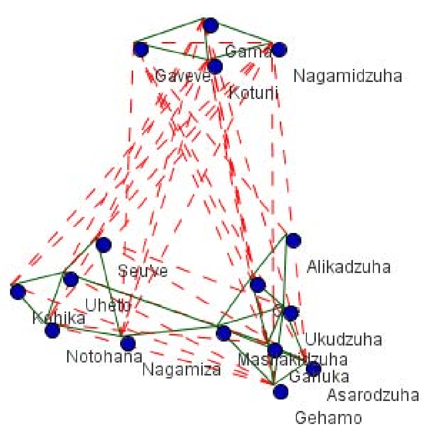
MovieLens 100k dataset consists of $100,000$ ratings on 1682 movies by 943 users. Ratings are in a 5-star scale. We establish a positive link between a user and an item if the user has rated the item with a score at least 0.2 larger than her average rating. In the same way, a negative edge is put when the corresponding rating is 0.2 smaller than the user's average rating. The MovieLens layouts are provided in Figure 6.



(a) Epinions layout. A cluster is detectable lower than the center, slightly to the left.

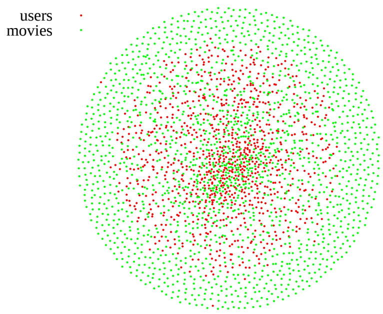


(b) Slashdot users. $k_1 = 19$, $k_2 = 10$, $k_3 = 10$.

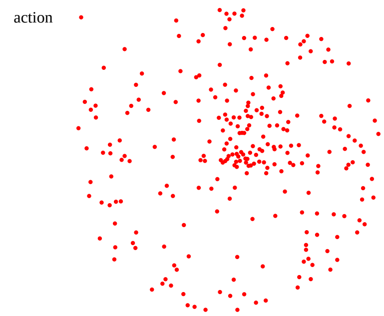



(c) The tribal groups of Eastern Central Highlands of New Guinea. Three higher order clusters described in [8] are revealed. Positive edges are in green. Negative edges are plot by dashed red lines.

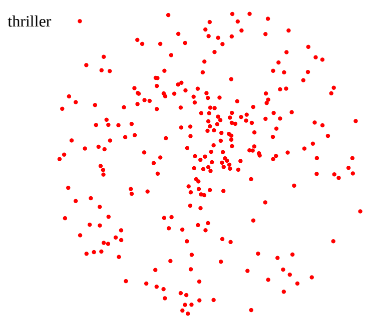
Figure 5: Epinions and Slashdot Signed LinLog layouts. The users having received a lot of positive votes are in the center. Zooming on the layouts reveals many lower order clusters. $k_1 = 19$, $k_2 = 10$, $k_3 = 10$. In the layout of the tribal groups of Eastern Central Highlands of New Guinea, $k_1 = 3$, $k_2 = 2$, $k_3 = 2$.



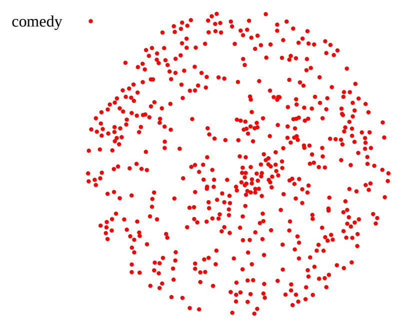

(a) Users and movies



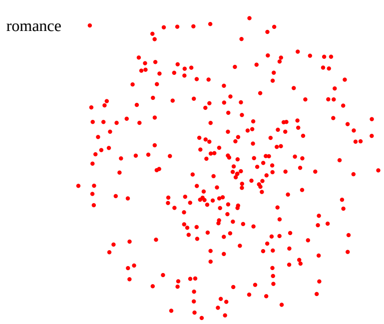

(b) Action



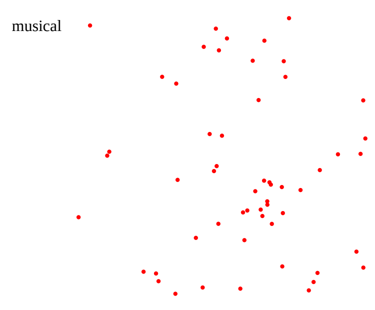

(c) Thriller




(d) Comedy




(e) Romance




(f) Musical

Figure 6: Signed LinLog Layouts of MovieLens 100k ratings. Active users and popular movies are in the center. Notice the changes in the density of movies. Some areas in the layout are pretty dense while some others are almost empty. Users lie close to the movies they have liked. $k_1 = 19$, $k_2 = 10$, $k_3 = 10$. In large sparse graphs, higher constants are more beneficial.

## 6 Two Other Clustering Signed Energy Models

In this section we present two other classes of energy models whose clustering version can cluster graphs in terms of different cuts. In contrast to the model of Section 4, they are not constructed by adding the model to its dual.

### 6.1 Two Clustering Models

The first one is defines as:

$$\begin{aligned} U_{SignedGraph} &= \sum_{\{u,v\}\in E^+} f(\|p_u - p_v\|) - \sum_{\{u,v\}\in E^-} f(\|p_u - p_v\|) + \\ & \sum_{\{u,v\}\in V^{(2)}} f(\|p_u - p_v\|) + g(\|p_u - p_v\|). \end{aligned}$$

In this model both negative and positive forces follow the same function. Like our previous model, this one is also capable of drawing disconnected graphs. The only case where the model does not converge is when there exist a subgraph whose each vertex has negative edges to all other vertices of the rest of the graph, i.e. $density^+(V_1, V_2) = 1$. Ignoring this special case that almost never happens in reality, we can state that the model can draw connected and disconnected graphs. In its clustering form it is written in the form of:

$$\begin{aligned} U_{SignedClustering} &= \sum_{\{u,v\}\in E^+} \|p_u - p_v\| - \sum_{\{u,v\}\in E^-} \|p_u - p_v\| + \\ & \sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| - \ln \|p_u - p_v\|. \end{aligned}$$

Here, we state the similar theorems about the clustering properties of this model. Proofs follow the same techniques as in Section 4.

**Theorem 6.1** *If $p^0$ is a drawing of a graph $G_s = (V, E)$ with minimum Signed Clustering energy then:*

$$\sum_{\{u,v\}\in E^+} \|p_u - p_v\| + \sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| = \sum_{\{u,v\}\in E^-} \|p_u - p_v\| + \left|V^{(2)}\right|.$$

**Theorem 6.2** *Let $p^0$ be a one-dimensional drawing of the signed graph $G_s = (V, E)$ with minimum Signed Clustering energy. Let $(V_1, V_2)$ be a partition of $V$ such that the vertices in $V_1$ have smaller positions than the vertices in $V_2$ (i.e. $\forall v_1 \in V_1,\ \forall v_2 \in V_2 : p_{v_1} < p_{v_2}$). Then:*

$$harm(V_1 \times V_2, p^0) = \frac{1}{1 + density^+(V_1, V_2) - density^-(V_1, V_2)}.$$

**Theorem 6.3** *If $p^0$ is a drawing of the signed graph $G_s = (V, E)$ with minimum Signed Clustering energy, $p^0$ also minimizes $\frac{arith^+(V^{(2)} \backslash E^-, p)}{geo(V^{(2)}, p)}$.*

The second class is defined as:

$$U_{SignedGraph} = \sum_{\{u,v\}\in E^+} f(\|p_u - p_v\|) + \sum_{\{u,v\}\in E^-} g(\|p_u - p_v\|) + \sum_{\{u,v\}\in V^{(2)}} g(\|p_u - p_v\|).$$

In this model, the set of positive edges acts as the backbone of the layout rendering convergence. Consequently, this model can not draw graphs disconnected by positive edges. Its clustering form is:

$$U_{SignedClustering} = \sum_{\{u,v\}\in E^+} \|p_u - p_v\| - \sum_{\{u,v\}\in E^-} \ln \|p_u - p_v\| - \sum_{\{u,v\}\in V^{(2)}} \ln \|p_u - p_v\|.$$

The clustering properties are as follows:

**Theorem 6.4** *If $p^0$ is a drawing of the positively connected signed graph $G_s = (V, E)$ with minimum Signed Clustering energy then:*

$$\sum_{\{u,v\}\in E^+} \|p_u - p_v\| = \left|E^-\right| + \left|V^{(2)}\right|.$$

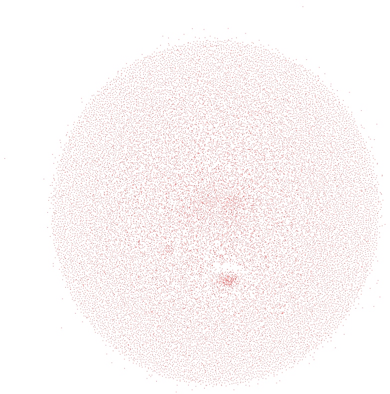
**Theorem 6.5** *Let $p^0$ be a one-dimensional drawing of the positively connected signed graph $G_s = (V, E)$ with minimum Signed Clustering energy. Let $(V_1, V_2)$ be a partition of $V$ such that the vertices in $V_1$ have smaller positions than the vertices in $V_2$ (i.e. $\forall v_1 \in V_1,\ \forall v_2 \in V_2 : p_{v_1} < p_{v_2}$). Then:*

$$harm^-(V_1 \times V_2, p^0) = \frac{1 + density^-(V_1, V_2)}{density^+(V_1, V_2)}.$$

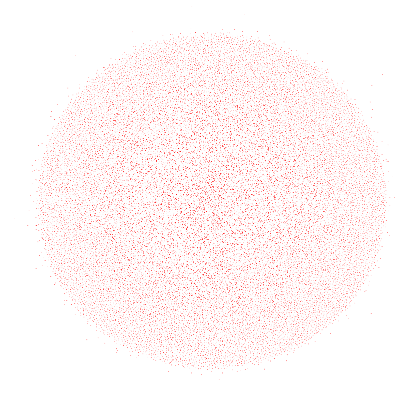
**Theorem 6.6** *If $p^0$ is a drawing of the positively connected signed graph $G_s = (V, E)$ with minimum Signed Clustering energy, $p^0$ also minimizes $\frac{arith(E^+)}{geo^-(V^{(2)},p)}$.*

## 6.2 Example Layouts

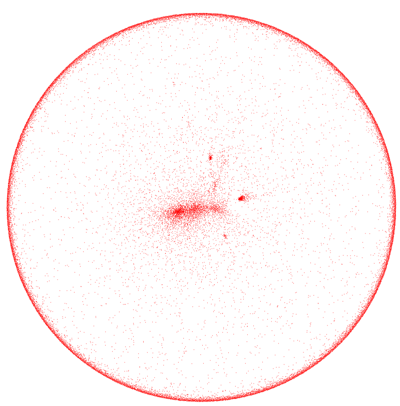
The real world layouts of these models are seen in Figures 7 and 8.



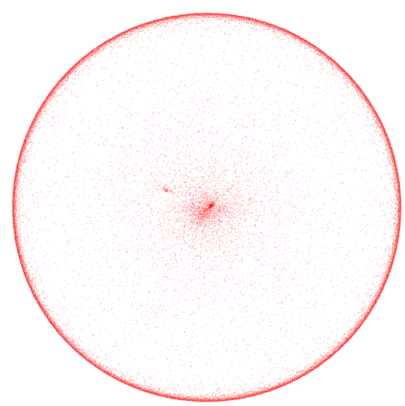
(a) Epinions (first model)



(b) Slashdot (first model)

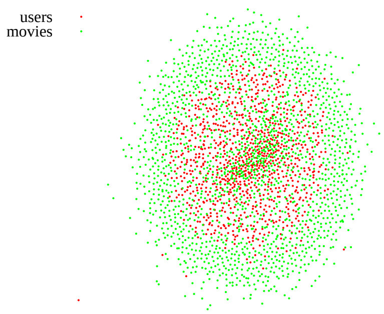


(c) Epinions (second model)

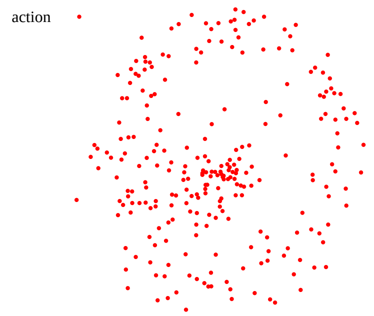


(d) Slashdot (second model)

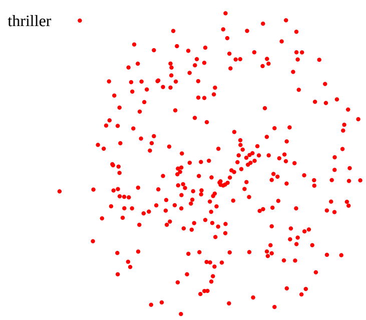
Figure 7: Epinions and Slashdot layouts with the first model ($k_1 = 19$, $k_2 = 10$, $k_3 = 10$) and the second model ($k_1 = 2$, $k_2 = 1$, $k_3 = 1$).



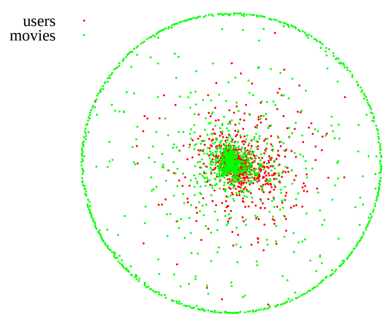

(a) Users and movies - first model



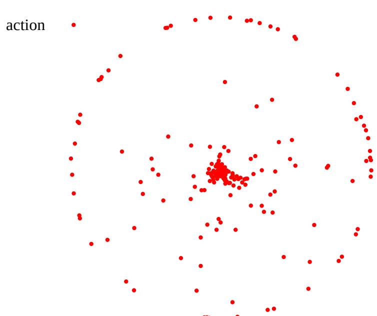

(b) Action movies - first model



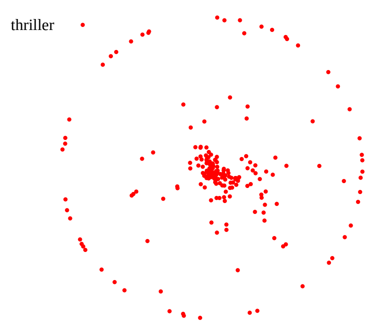

(c) Thriller movies - first model




(d) Users and movies - second model




(e) Action movies - second model




(f) Thriller movies - second model

Figure 8: MovieLens 100k ratings visualization with the first and second model, $k_1 = 19$, $k_2 = 10$, $k_3 = 10$.

# 7 Energy Models for Drawing Bipartite Graphs

The user/item interaction can be modeled by bipartite graphs in collaborative rating networks. The two partitions of such a graph represent users and items. There exists an edge between a user and an item if the user has done some activity on the item. This is in the form of explicit rating in its less noisiest form, but may also refer to implicit data like clicking history. Real-world examples of collaborative filtering networks are MovieLens and Netflix where users rate movies in a 5-star scale to express their opinion on them. Visualization of these networks leads to detection of user and item clusters, being in turn a helpful asset to recommendation systems and market analysis.

Formally, we show bipartite graphs with $G_b = (V_1, V_2, E_b)$, where $E_b \subset V_1 \times V_2$ is the edge set, and $V_1$, $V_2$ are two disjoint sets of vertices representing the vertex partitions of the graph. Bipartite graphs are different from unipartite graphs in that there is no internal edge in the two partitions, and edges only exist between vertices belonging to two different partitions. Consequently, efficient bipartite graph visualization necessitates proper models. Standard graph visualization tools and libraries (e.g. NodeXL for Excel or Jung library for Java) do not offer such models. In this section we suggest an abstract equation for bipartite energy models. Then, we derive the clustering properties of Bipartite LinLog model. We define the bipartite density between two clusters $V_l$ and $V_r$ of a bipartite graph, where each cluster contains vertices from both $V_1$ and $V_2$, as:

$$density^b_{lr} = \frac{|E_{lr}|}{|V_{1l}|\,|V_{2r}| + |V_{1r}|\,|V_{2l}|}. \tag{3}$$

The main problem of Equation (4) for visualizing bipartite graphs is that the neutral energy $g(\|p_u - p_v\|)$ exists between every two vertices including $(u, v):\ u \in V_1,\ v \in V_2$. For bipartite graphs, this leads in each partition influencing the configuration of the other out of the interaction set fully stated by edges. In general, the two partitions represent two totally different entities (e.g. users vs. items). Therefore, it is desirable to omit any mutual influence between them unless with the intermediate of edges. Having this idea in mind, we arrive at the following definition for bipartite energy models:

$$U_{Bipartite} = \sum_{\{u,v\} \in E_b} f(\|p_u - p_v\|) + \sum_{\{u,v\} \in V_1^{(2)} \cup V_2^{(2)}} g(\|p_u - p_v\|).$$

The bipartite version of Signed LinLog is defined as:

$$U_{BipartiteLinLog} = \sum_{\{u,v\} \in E_b} \|p_u - p_v\| - \sum_{\{u,v\} \in V_1^{(2)} \cup V_2^{(2)}} \ln \|p_u - p_v\|.$$

Theorem 7.1 states that in a layout with minimum Bipartite LinLog energy, the total length of edges is fixed; it is equal to the sum of squares of the cardinality of the two partitions.

**Theorem 7.1** *If* $p_0$ *is a drawing of a bipartite graph* $G_b = (V_1, V_2, E_b)$ *with minimum Bipartite LinLog energy, then:*

$$\sum_{\{u,v\}\in E_b} \|p_u - p_v\| = |V_1|^2 + |V_2|^2 .$$

**Theorem 7.2** *Let* $p^0$ *be a one-dimensional layout of the bipartite graph* $G_b = (V_1,\ V_2, E_b)$*. Let* $V_{1l}, V_{1r}$ *be a partition of* $V_1$*, and* $V_{2l}$*,* $V_{2r}$ *a partition of* $V_2$ *such that all vertices in* $V_{1l}$ *and* $V_{2l}$ *have smaller position than vertices in* $V_{1r}$ *and* $V_{2r}$ *(i.e.* $\forall v_l \in V_{1l} \cup V_{2l}$*,* $\forall v_r \in V_{2r} \cup V_{1r}$ *:* $p_{v_l} < p_{v_r}$*). Then:*

$$harm(V_{1l} \times V_{1r} \cup V_{2l} \times V_{2r}) = \frac{|V_{1l} \times V_{1r}| + |V_{2l} \times V_{2r}|}{|E_{lr}|} = \frac{1}{density^b_{lr}},$$

*where* $E_{lr} = E_{V1l \times V_{2r}} \cup E_{V2l \times V_{1r}}$*, i.e. edges connecting the vertices on the left half-line to the vertices of the right half-line.*

**Theorem 7.3** *Let* $G_b = (V_1, V_2, E_b)$ *be a connected bipartite graph, and let* $p^0$ *be a layout of* $G_b$ *with minimum Bipartite LinLog energy. Then* $p^0$ *is a layout of* $G_b$ *that minimizes* $\frac{arith(E_b,p)}{geo(V_1^{(2)} \cup V_2^{(2)},p)}$.

# 8 Energy Models for Drawing Signed Bipartite Graphs

We represent bipartite signed graphs with $G_{bs} = (V_1, V_2, E_{bs})$, where signed edge set is represented by $E_{bs} = E^+_{bs} \cup E^-_{bs}$, $E^+_{bs}, E^-_{bs} \subset V_1 \times V_2$, $E^+_{bs} \cap E^-_{bs} = \emptyset$, and $V_1$ and $V_2$ denote the two vertex partitions. We define the positive and negative density for bipartite signed graphs as follows:

$$density^{b,+}_{lr} = \frac{\left|E^+_{lr}\right|}{|V_{1l}|\,|V_{2r}| + |V_{1r}|\,|V_{2l}|}, \quad density^{b,-}_{lr} = \frac{\left|E^-_{lr}\right|}{|V_{1l}|\,|V_{2r}| + |V_{1r}|\,|V_{2l}|}.$$

The abstract form of bipartite signed energy models is defined as:

$$\begin{aligned} U_{BipartiteSignedGraph} \quad = & \sum_{\{u,v\}\in E^+_{bs}} f(\|p_u - p_v\|) + \sum_{\{u,v\}\in E^-_{bs}} g(\|p_u - p_v\|) + \\ & \sum_{\{u,v\}\in V_1^{(2)} \cup V_2^{(2)}} f(\|p_u - p_v\|) + g(\|p_u - p_v\|). \end{aligned}$$

Bipartite Signed LinLog model is defined as:

$$\begin{aligned} U_{BipartiteSignedLinLog} \quad = & \sum_{\{u,v\}\in E^+_{bs}} \|p_u - p_v\| - \sum_{\{u,v\}\in E^-_{bs}} \ln\|p_u - p_v\| + \\ & \sum_{\{u,v\}\in V_1^{(2)} \cup V_2^{(2)}} \|p_u - p_v\| - \ln\|p_u - p_v\|. \end{aligned}$$

The theorems of prior sections may be easily generalized to signed bipartite graphs. The proofs are quite similar to previous ones. Here, we state the results:

**Theorem 8.1** *If $p_0$ is a drawing of a bipartite signed graph $G_{bs} = (V_1, V_2, E_{bs})$ with minimum Bipartite Signed LinLog energy, then:*

$$\sum_{\{u,v\} \in V_1^{(2)}} \|p_u - p_v\| + \sum_{\{u,v\} \in V_2^{(2)}} \|p_u - p_v\| + \sum_{E_b^+} \|p_u - p_v\| = |V_1|^2 + |V_2|^2 + \left|E_b^-\right|.$$

**Theorem 8.2** *Let $p^0$ be a one-dimensional layout of the bipartite signed graph $G_{bs} = (V_1, V_2, E_{bs})$. Let $V_{1l}, V_{1r}$ be a partition of $V_1$, and $V_{2l}$, $V_{2r}$ a partition of $V_2$ such that all vertices in $V_{1l}$ and $V_{2l}$ have smaller position than vertices in $V_{1r}$ and $V_{2r}$ (i.e. $\forall v_l \in V_{1l} \cup V_{2l}$, $\forall v_r \in V_{2r} \cup V_{1r} : p_{v_l} < p_{v_r}$). Then:*

$$harm(V_{1l} \times V_{1r} \cup V_{2l} \times V_{2r} \cup E_{bs}^-) = \frac{\kappa + density_{lr}^{b,-}}{\kappa + density_{lr}^{b,+}},$$

where

$$\kappa = \frac{|V_{1l}|\,|V_{1r}| + |V_{2l}|\,|V_{2r}|}{|V_{1l}|\,|V_{2r}| + |V_{1r}|\,|V_{2l}|}$$

**Theorem 8.3** *Let $G_{bs} = (V_1, V_2, E_{bs})$ be a connected bipartite signed graph, and let $p^0$ be a layout of $G_{bs}$ with minimum Bipartite Signed LinLog energy. Then $p^0$ is a layout of $G_{bs}$ that minimizes $\frac{arith(V_1^{(2)} \cup V_2^{(2)} \cup E_{bs}^+, p)}{geo(V_1^{(2)} \cup V_2^{(2)} \cup E_{bs}^-, p)}$.*

# 9 Conclusion

The problem of signed graph visualization has been ignored so far. In this paper, we suggested energy models for drawing signed graphs. We focused on preserving the community structure inside sociograms. A dual model was first suggested to draw graphs containing uniquely repulsive edges. An abstract equation for signed energy models was consequently proposed by adding up the primary models by their dual. Signed LinLog was presented as a model preserving the community structure inside graphs. We analytically derived its clustering properties, and argued the effect of force constants. The most important results were as follows. Signed LinLog draws positive edges short, and negative edges long w.r.t the mean neutral distance between vertices. This property renders the revelation of clusters. Signed LinLog generates layouts of graph bipartitions where the negative harmonic mean distance between the partitions increases with the positive density, and decreases with the negative density. Our signed energy models are capable of drawing disconnected graphs. Disconnected vertices are plot in places with distances as uniform as possible from the rest of the vertices.

# References


[1] J. Barnes and P. Hut. A hierarchical $o(nlogn)$ force-calculation algorithm. In *Nature*, pages 446–449, 1986.

[2] U. Brandes. Drawing on physical analogies. In M. Kaufmann and D. Wagner, editors, *Drawing Graphs*, volume 2025, pages 71–86. Springer, 2001.

[3] F. Dabek, R. Cox, F. Kaashoek, and R. Morris. Vivaldi: a decentralized network coordinate system. In *Applications, technologies, architectures, and protocols for computer communications*, pages 15–26, 2004.

[4] R. Davidson and D. Harel. Drawing graphs nicely using simulated annealing. In *ACM Transactions on Graphics*, pages 301–331, 1996.

[5] P. Eades. A heuristic for graph drawing. In *Congressus Numerantium*, pages 149–160, 1984.

[6] T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. In *Software-Practice and Experience*, pages 1129–1164, 1991.

[7] GroupLens. *MovieLens Datasets*, 2010. `http://www.grouplens.org/node/73#attachments`.

[8] P. Hage and F. Harary. *Structural Models in Anthro-pology.* Cambridge University Press, 1983.

[9] K. M. Hall. An r-dimensional quadratic placement algorithm. *Management Science*, pages 219–229, 1970.

[10] F. Heider. Attitudes and cognitive organization. *The Journal of Psychology*, pages 107–112, 1946.

[11] T. Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.*, pages 7–15, 1989.

[12] Y. Koren, L. Carmel, and D. Harel. Ace: a fast multiscale eigenvectors computation for drawing huge graphs. In *INFOVIS*, pages 137–144, 2002.

[13] Y. Koren and A. Çivril. The binary stress model for graph drawing. In *Graph Drawing*, pages 193–205, 2009.

[14] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. D. Luca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *SDM*, pages 559–570, 2010.

[15] S. Milgram. The small world problem. *Psychology Today 1*, 1967.

[16] M. E. J. Newman. The structure and function of complex networks. *SIAM REVIEW*, pages 167–256, 2003.

[17] M. E. J. Newman. Analysis of weighted networks. *Physical Review E*, 2004.

[18] A. Noack. Energy models for drawing clustered small-world graphs. Technical report, Dep. of Comp. Sience, Brandenburg Tech. University at Cottbus, 2003.

[19] A. Noack. Energy models for graph clustering. *Journal of Graph Algorithms and Applications*, pages 453–480, 2007.

[20] K. E. Read. Cultures of the central highlands, new guinea. *SouthWestern Journal of Anthropology*, 1954.

[21] SNAP. *Stanford Large Network Dataset Collection*, 2010. `http://snap.stanford.edu/data`.

[22] D. Tunkelang. Jiggle: Java interactive graph layout environment. In *Graph Drawing (GD)*, pages 413–422, 1998.

## A Theorems and Proofs

**Proof of Theorem 4.3** We first suppose that the total distance between all vertices is fixed, and prove the theorem under this restriction. This restriction will be removed at the end of the proof using proof by contradiction. Suppose that $p^0$ is a drawing minimizing Dual LinLog energy, that is, it is a solution to the problem:

$$\text{minimize } U_{DualLinLog}(p) = \sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| - \sum_{\{u,v\}\in E^-} \ln \|p_u - p_v\|.$$

The total distance between all vertices in $p^0$ would be a positive value, say $c$:

$$\sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| = c.$$

Hence, $p^0$ is also a solution to:

$$\text{minimize } - \sum_{\{u,v\}\in E^-} \ln \|p_u - p_v\| \text{ subject to } arith(V^{(2)}, p) = \frac{c}{\left|V^{(2)}\right|}.$$

Since $\sqrt[\left|E^-\right|]{\exp(x)}$ is an increasing function of its argument, the above expression is equivalent to:

$$\text{minimize } \sqrt[\left|E^-\right|]{\exp(- \sum_{\{u,v\}\in E^-} \ln \|p_u - p_v\|)} \text{ subject to } arith(V^{(2)}, p) = \frac{c}{\left|V^{(2)}\right|},$$

that is, minimize $\frac{1}{geo(E^-,p)}$ subject to $arith(V^{(2)}, p) = \frac{c}{\left|V^{(2)}\right|}$. If we multiply this result by $arith(V^{(2)}, p)$, we have:

$$\text{minimize } \frac{arith(V^{(2)}, p)}{geo(E^-, p)} \text{ subject to } arith(V^{(2)}, p) = \frac{c}{\left|V^{(2)}\right|}.$$

We now remove the restriction. Suppose there exists a drawing $q^0$ for which $\frac{arith(V^{(2)},q^0)}{geo(E^-,q^0)} < \frac{arith(V^{(2)},p^0)}{geo(E^-,p^0)}$. We can define a scaling $q^1 = \frac{c}{\left|V^{(2)}\right| arith(V^{(2)},q^0)} q^0$ for which $arith(V^{(2)}, q^1) = \frac{c}{\left|V^{(2)}\right|}$ and $\frac{arith(V^{(2)},q^1)}{geo(E^-,q^1)} = \frac{arith(V^{(2)},q^0)}{geo(E^-,q^0)} < \frac{arith(V^{(2)},p^0)}{geo(E^-,p^0)}$. This contradicts the previous results. Hence, $q^0$ does not exist, and we can always remove the restriction. ∎

**Proof of Theorem 4.6** Suppose that $p^0$ is a one-dimensional layout minimizing the Dual LinLog energy. Then, adding $d \in R$ to the coordinates of all vertices in $V_1$ in a way that they remain smaller than the coordinates of all vertices in $V_2$ (i.e. $\max\{p_u|u \in V_1\} < \min\{p_u|u \in V_2\}$), the Dual LinLog energy is:

$$U_{DualLinLog}(d, p^0) = \sum_{\{u,v\} \in V_1^{(2)} \cup V_2^{(2)}} (\|p_u - p_v\|) - \sum_{\{u,v\} \in E^-_{V_1^{(2)}} \cup E^-_{V_2^{(2)}}} \ln(\|p_u - p_v\|)$$
$$\sum_{\{u,v\} \in V_1 \times V_2} (\|p_u - p_v\| + d) - \sum_{\{u,v\} \in E^-_{V_1 \times V_2}} \ln(\|p_u - p_v\| + d).$$

This function must have a global minimum at $d = 0$ because $p$ is a layout with minimum Dual LinLog Energy:

$$\begin{aligned} U'_{DualLinLog}(d, p^0) &= |V_1|\,|V_2| - \sum_{\{u,v\} \in E^-_{V_1 \times V_2}} \frac{1}{\|p_u - p_v\| + d}. \\ U'_{DualLinLog}(d = 0, p^0) &= |V_1|\,|V_2| - \sum_{\{u,v\} \in E^-_{V_1 \times V_2}} \frac{1}{\|p_u - p_v\|} = 0. \\ |V_1|\,|V_2| &= \frac{\left|E^-_{V_1 \times V_2}\right|}{harm(E^-_{V_1 \times V_2}, p)}. \end{aligned}$$

∎

**Theorem A.1 ([18])** *If $p^0$ is a solution to $Minimize\ U_{LinLog}$, then:*

$$\sum_{\{u,v\} \in E^+} \|p_u - p_v\| = \left|V^{(2)}\right|.$$

Theorem A.1 states that the total length of edges is always fixed in layouts with minimum LinLog energy. The value of this total length is the cardinality of $V^{(2)}$.

**Theorem A.2** *If $p^0$ is a solution to $Minimize\ U_{DualLinLog}$, then:*

$$\sum_{\{u,v\} \in V^{(2)}} \|p_u - p_v\| = \left|E^-\right|.$$

**Proof** Suppose $p^0$ is a drawing with minimum Dual LinLog energy. If we multiply all coordinates by some $d \in R$, the energy of the system becomes:

$$U_{DualLinLog}(d, p^0) = \sum_{\{u,v\} \in V^{(2)}} d\|p_u - p_v\| - \sum_{\{u,v\} \in E^-} \ln d\|p_u - p_v\|.$$

Since $p^0$ is a drawing with minimum Dual LinLog energy every scaling of $p^0$ must lead to an increase in the energy of the system. Hence, the above expression has a minimum at $d = 1$.

$$\begin{aligned} U'_{DualLinLog}(d, p^0) &= \sum_{\{u,v\} \in V^{(2)}} \|p_u - p_v\| - \frac{\left|E^-\right|}{d}, \\ U'_{DualLinLog}(d = 1, p^0) &= \sum_{\{u,v\} \in V^{(2)}} \|p_u - p_v\| - \left|E^-\right| = 0. \end{aligned}$$

∎

Theorem A.2 states that the total Euclidean distance of all vertices is always fixed in layouts with minimum Dual LinLog energy. It is equal to the number of edges.

**Theorem A.3** *If $p^0$ is a drawing of a graph $G_s(V, E)$ with minimum Signed LinLog energy then:*

$$\sum_{\{u,v\}\in E^+} \|p_u - p_v\| + \sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| = \left|E^-\right| + \left|V^{(2)}\right|.$$

**Proof** Suppose $p^0$ is a drawing with minimum Signed LinLog energy. If we multiply all coordinates in $p^0$ by $d \in R$, the energy of the system is:

$$\begin{aligned} U_{SignedLinLog}(d, p^0) = & \sum_{\{u,v\}\in E^+} d\|p_u - p_v\| - \sum_{\{u,v\}\in E^-} \ln d\|p_u - p_v\| \\ & + \sum_{\{u,v\}\in V^{(2)}} d\|p_u - p_v\| - \ln d\|p_u - p_v\|. \end{aligned}$$

Since $p^0$ is a drawing with minimum energy, this equation has a minimum at $d = 1$, that is:

$$\begin{aligned} U'_{SignedLinLog}(d, p^0) &= \sum_{\{u,v\}\in E^+} \|p_u - p_v\| - \frac{\left|E^-\right|}{d} + \sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| - \frac{\left|V^{(2)}\right|}{d}, \\ U'_{SignedLinLog}(d = 1, p^0) &= \sum_{\{u,v\}\in E^+} \|p_u - p_v\| - \left|E^-\right| + \sum_{\{u,v\}\in V^{(2)}} \|p_u - p_v\| - \left|V^{(2)}\right| = 0. \end{aligned}$$

∎

Theorem A.3 states that the sum of the total length of positive edges with the total Euclidean distance between all vertices is always a fixed value in the drawings with minimum Signed LinLog energy.

# B Distance Interpretability in 1+ Dimensional Signed LinLog Bipartition Layouts

In this appendix we explain the approximate generalizability of Theorem 4.4 to 1+ dimensions. The following theorem holds exactly for layouts with any number of dimensions:

**Theorem B.1** *Let $p$ be a $D$-dimensional drawing of $G_s = (V, E)$ with minimum Weighted Signed LinLog energy. Let $(S_1, S_2)$ be a bipartition of the drawing space by any hyperplane $H$ defined by $\sum_{i\in I} a_i x_i = b$, $I \in \binom{\{1,\cdots,D\}}{D-1}$. If $(V_1, V_2)$ is a bipartition of vertices in a way that $\forall u \in V_1 : p_u \in S_1$, and $\forall v \in V_2 : p_v \in S_2$, that is $\forall u \in V_1 : \sum_i a_i x_i^u < b$ and $\forall v \in V_2 : \sum_i a_i x_i^v > b$,* *then:*

$$\sum_{\{u,v\}\in E^+_{V_1\times V_2}} k_1 \frac{\sum_{i=1}^D (x_i^u - x_i^v)}{\|p_u - p_v\|} + \sum_{\{u,v\}\in V_1\times V_2} \frac{\sum_{i=1}^D (x_i^u - x_i^v)}{\|p_u - p_v\|} =$$

$$\sum_{\{u,v\}\in E^-_{V_1\times V_2}} k_2 \frac{\sum_{i=1}^D (x_i^u - x_i^v)}{\|p_u - p_v\|^2} + \sum_{\{u,v\}\in V_1\times V_2} k_3 \frac{\sum_{i=1}^D (x_i^u - x_i^v)}{\|p_u - p_v\|^2}.$$

**Proof** If we add some distance vector $\vec{d} = (d_1, \ldots, d_D)$ to all vertices in $V_1$ in a way that none of them enter $S_2$, i.e. $\forall u \in V_1 : \sum_i a_i(x_i^u + d_i) < b$, the energy of the new drawing is:

$$\begin{aligned}
U^{new} = & \sum_{\{u,v\}\in E^+_{V_1^{(2)}} \cup E^+_{V_2^{(2)}}} k_1 \|p_u - p_v\| - \sum_{\{u,v\}\in E^-_{V_1^{(2)}} \cup E^-_{V_2^{(2)}}} k_2 \ln \|p_u - p_v\| + \\
& \sum_{\{u,v\}\in V_1^2 \cup V_2^2} (\|p_u - p_v\| - k_3 \ln \|p_u - p_v\|) + \sum_{\{u,v\}\in E^+_{V_1\times V_2}} k_1 \sqrt{\sum_{i=1}^D (x_i^u - x_i^v + d_i)^2} - \\
& \sum_{\{u,v\}\in E^-_{V_1\times V_2}} k_2 \ln \sqrt{\sum_{i=1}^D (x_i^u - x_i^v + d_i)^2} + \\
& \sum_{\{u,v\}\in V_1\times V_2} (\sqrt{\sum_{i=1}^D (x_i^u - x_i^v + d_i)^2} - k_3 \ln \sqrt{\sum_{i=1}^D (x_i^u - x_i^v + d_i)^2}).
\end{aligned}$$

The partial derivative of this function with respect to $d_i$ is:

$$\begin{aligned}
\frac{\partial U_{new}}{\partial d_i} = & \sum_{\{u,v\}\in E^+_{V_1\times V_2}} k_1 \frac{x_i^u - x_i^v + d_i}{\sqrt{\sum_{i=1}^D (x_i^u - x_i^v + d_i)^2}} - \sum_{\{u,v\}\in E^-_{V_1\times V_2}} k_2 \frac{x_i^u - x_i^v + d_i}{\sum_{i=1}^D (x_i^u - x_i^v + d_i)^2} + \\
& \sum_{\{u,v\}\in V_1\times V_2} (\frac{x_i^u - x_i^v + d_i}{\sqrt{\sum_{i=1}^D (x_i^u - x_i^v + d_i)^2}} - k_3 \frac{x_i^u - x_i^v + d_i}{\sum_{i=1}^D (x_i^u - x_i^v + d_i)^2}).
\end{aligned}$$

Since $p$ is a layout with minimum Signed LinLog energy, the application of any non zero vector $\vec{d}$ must result in increase of energy. Hence, the gradient vector of $U_{new}$ must be zero when $\vec{d} = 0$, that is $\forall d_i : \frac{\partial U_{new}}{\partial d_i} = 0$. Hence $\sum_{i=1}^D \frac{\partial U_{new}}{\partial d_i} = 0$.

$$\begin{aligned}
\sum_{i=1}^D \frac{\partial U_{new}}{\partial d_i} = & \sum_{\{u,v\}\in E^+_{V_1\times V_2}} k_1 \frac{\sum_{i=1}^D (x_i^u - x_i^v)}{\sqrt{\sum_{i=1}^D (x_i^u - x_i^v)^2}} - \sum_{\{u,v\}\in E^-_{V_1\times V_2}} k_2 \frac{\sum_{i=1}^D (x_i^u - x_i^v)}{\sum_{i=1}^D (x_i^u - x_i^v)^2} + \\
& \sum_{\{u,v\}\in V_1\times V_2} (\frac{\sum_{i=1}^D (x_i^u - x_i^v)}{\sqrt{\sum_{i=1}^D (x_i^u - x_i^v)^2}} - k_3 \frac{\sum_{i=1}^D (x_i^u - x_i^v)}{(x_i^u - x_i^v)^2}) = 0.
\end{aligned}$$

∎

For D-dimensional layouts of graphs clusterizable to some extent, Theorem B.1 leads to the following useful corollary:

**Corollary B.1** *Let $p$ be a $D$-dimensional drawing of $G_s = (V, E)$ with minimum Weighted Signed LinLog energy. For any non-scaling linear transformation of the coordinate system*[3] *that partitions the vertices into $(V_1, V_2)$ in a way that in the new coordinate system $\forall u \in V_1, v \in V_2, 1 \leq i \leq D : x_i^u < x_i^v$:* [4]

$$\sum_{\{u,v\} \in E^+_{V_1 \times V_2}} k_1 \frac{\|p_u - p_v\|_{Man}}{\|p_u - p_v\|} + \sum_{\{u,v\} \in V_1 \times V_2} \frac{\|p_u - p_v\|_{Man}}{\|p_u - p_v\|} =$$

$$\sum_{\{u,v\} \in E^-_{V_1 \times V_2}} k_2 \frac{\|p_u - p_v\|_{Man}}{\|p_u - p_v\|^2} + \sum_{\{u,v\} \in V_1 \times V_2} k_3 \frac{\|p_u - p_v\|_{Man}}{\|p_u - p_v\|^2},$$

*where $\|p_u - p_v\|_{Man} = \sum_{i=1}^{D} \left|x_i^u - x_i^v\right|$ is the* Manhatan distance *between $p_u$ and $p_v$.*

We know from Corollary 4.1 that force constants may be adjusted to shorten positive edges and lengthen negative edges as much as necessary. Hence, *provided the graph is clusterizable into two convex subgraphs*, we may modify the constants to decrease the diameter of clusters (i.e. the maximum Euclidean distance between pairs of a cluster) and increase their distance as much as desired. If the vertices are concentrated and far from each other, the Euclidean and Manhatan distance become almost equal. In this case we can state:

**Corollary B.2** *Let $p$ be a $D$-dimensional drawing of $G_s = (V, E)$ with minimum Weighted Signed LinLog energy. If a bipartition of vertices $(V_1, V_2)$ exists in a way that the diameter of $V_1$ and $V_2$ is small compared to their distance, then:*

$$harm^{k_2-, k_3 d}(V_1 \times V_2, p^0) \approx \frac{k_3 + k_2 density^-(V_1, V_2)}{1 + k_1 density^+(V_1, V_2)}.$$

**Proof** Putting $\|p_u - p_v\| \approx \|p_u - p_v\|_{Man}$ into Theorem B.1, we obtain:

$$k_1 \left|E^+_{V_1 \times V_2}\right| + |V_1 \times V_2| = \sum_{\{u,v\} \in E^-_{V_1 \times V_2}} \frac{k_2}{\|p_u - p_v\|} + \sum_{\{u,v\} \in V_1 \times V_2} \frac{k_3}{\|p_u - p_v\|}.$$

Replacing the right side by $\frac{k_2 \left|E^-_{V_1 \times V_2}\right| + k_3 |V_1 \times V_2|}{harm^{k_2-, k_3 d}(V_1 \times V_2, p^0)}$, where

$$\begin{aligned} harm^{k_2-, k_3 d}(V_1 \times V_2, p^0) &= \frac{k_2 \left|E^-_{V_1 \times V_2}\right| + k_3 \left|V_1 \times V_2\right|}{\sum_{\{u,v\} \in V_1 \times V_2} \frac{\lambda'_{uv}}{\|p_u - p_v\|}}, \quad \lambda'_{uv} = \begin{cases} k_2 + k_3 & if\ \{u,v\} \in E^-_{V_1 \times V_2} \\ k_3 & if\ \{u,v\} \in V_1 \times V_2 \backslash E^-_{V_1 \times V_2} \end{cases} \\ &= \frac{k_2 \left|E^-_{V_1 \times V_2}\right| + k_3 \left|V_1 \times V_2\right|}{\frac{k_2 \left|E^-_{V_1 \times V_2}\right|}{harm(E^-_{V_1 \times V_2})} + \frac{k_3 |V_1 \times V_2|}{harm(V_1 \times V_2, p)}}. \end{aligned}$$

the result is derived. ∎

[3] This causes no change to the energy of the system.

[4] Notice such transformation does not exist for the layouts of all graphs.