Forming Compatible Teams in Signed Networks

Ioannis Kouvatis 1, Konstantinos Semertzidis 2
Maria Zerva 1, Evaggelia Pitoura 1, Tsaparas Panayiotis 1

1 Department of Computer Science and Engineering, University of Ioannina, Greece
2 IBM Research, Dublin, Ireland


Abstract

The problem of team formation in a social network asks for a set of individuals who not only have the required skills to perform a task but who can also communicate effectively with each other. Existing work assumes that all links in a social network are positive, that is, they indicate friendship or collaboration between individuals. However, it is often the case that the network is signed, that is, it contains both positive and negative links, corresponding to friend and foe relationships. Building on the concept of structural balance, we provide definitions of compatibility between pairs of users in a signed network, and algorithms for computing it. We then define the team formation problem in signed networks, where we ask for a compatible team of individuals that can perform a task with small communication cost. We show that the problem is NP-hard even when there are no communication cost constraints, and we provide heuristic algorithms for solving it. We present experimental results to investigate the properties of the different compatibility definitions, and the effectiveness of our algorithms.

1 Introduction

Given a task that requires a set of skills, a pool of workers who possess some of the skills and are organized in a social network, team formation refers to finding a subset of the workers that collectively cover the skills and can communicate effectively with each other [9]. The communication cost of the team is measured using the distances between the team members in the network. The idea is that socially well connected users will be more effective in working together.

Since the pioneering work in [9], there has been considerable research activity on the problem [1, 8, 12]. All existing work assumes that the social network contains only positive ties between individuals. That is, the edges denote friendship or successful collaboration between two users. However, very often, we have signed networks with both positive and negative ties. A negative edge indicates a contentious relationship and inability of two users to collaborate and thus they should not be in the same team.

In this paper, we study the problem of team formation in signed networks. In addition to the known requirements of the team formation problem, we ask that the team contains users that are all compatible with each other. Defining compatibility in a signed network is an interesting problem in itself. Clearly, users connected with a positive edge are compatible, while users connected with a negative edge are incompatible. We infer the compatibility of non-connected pairs of users by using the structure of the graph, and the principle of structural balance. Structural balance [4] is based on the premise that “the friend of my friend is my friend”, “the enemy of my enemy is my friend”, and “the enemy of my friend is my enemy”. Using this premise, we determine the compatibility of two users by looking at the paths that connect them. For example, a path of two positive or two negative edges indicates compatibility, while a path of one positive and one negative edge indicates incompatibility. We formalize this idea, and we provide definitions of compatibility of varying strictness. We perform experiments with four real datasets, where we evaluate the different compatibility definitions, and the performance of the algorithms for the team formation problem.

2 Problem Definition

We are given as input a pool of n individuals organized in an undirected signed graph $G = (V, E)$. Each node in $V$ corresponds to an individual, and $E = \{(u,v) \in V \times V : \ell \in \{+1, -1\}\}$ is a set of edges labeled as either positive or negative to indicate that $u$ and $v$ are friends or enemies respectively. We assume that $G$ is connected. We will use a function $\text{sign} : E \rightarrow \{+1, -1\}$ that returns the label of each edge in $E$.

We are also given as input a universe $S$ of skills. Each individual $u$ in $V$ possesses a set of skills, $\text{skill}(u) \subseteq S$. We define a task as the subset of skills $T \subseteq S$ required for its completion. Given a task, the team formation problem asks for a team $X$ of individuals, $X \subseteq V$, that collectively covers the required skills and whose members can work together effectively. The effectiveness of a team is typically quantified by the communication cost, $\text{Cost}(X)$, of the team defined as some function of the distances of its members in the graph.

However, when the graph contains both positive and negative edges, we need to take into account that some individuals, although close in the graph, may not be compatible with each other. To capture whether two users are compatible, we introduce a relation $\text{Comp} \subseteq V \times V$ such that $(u,v) \in \text{Comp}$, if and only if, if $u$ and $v$ can work together. Two natural requirements for $\text{Comp}$ are (1) reflexivity: $(u,u) \in \text{Comp}$, and (2) symmetry: if $(u,v) \in \text{Comp}$, then $(v,u) \in \text{Comp}$. Furthermore, we require that the $\text{Comp}$ relation satisfies the following two intuitive properties:

1. **Positive Edge Compatibility**: For all $(u,v) \in E$, such that $\text{sign}(u,v) = +1$, $(u,v) \in \text{Comp}$.
2. **Negative Edge Incompatibility**: For all $(u,v) \in E$, such that $\text{sign}(u,v) = -1$, $(u,v) \notin \text{Comp}$.

We will provide various definitions of compatibility in Section 3. We now define formally our problem.

**Definition 2.1.** (Team Formation in Signed Networks (TFSN)) Given a signed graph $G = (V, E)$, a compatibility relation $\text{Comp}$, and a task $T$, find $X \subseteq V$ such that (1) $\cup_{u \in X} \text{skill}(u) \subseteq T$, (2) for each $u,v \in X$, $(u,v) \in \text{Comp}$, and (3) $\text{Cost}(X)$ is minimized.

The TFSN problem contains as a special case the original team formation problem which is NP-hard [9], thus TFSN is also NP-hard. Moreover, we have shown that just finding a set of
compatible users is NP-hard. Let TFSNC denote the simplified version of the TFSN problem, where we drop the third requirement of minimizing the cost. In particular, we have proven the following theorem.

**Theorem 2.2.** The decision version of TFSNC is NP-hard for any compatibility relation that satisfies positive edge compatibility and negative edge incompatibility.

### 3. USER COMPATIBILITY

We start with two basic definitions of compatibility.

**Definition 3.1.** Direct Positive Edge (DPE) compatibility: \( \text{Comp}_{DPE} = \{(u, v) \in V \times V : (u, v, +1) \in E\} \).

**Definition 3.2.** No Negative Edge (NNE) compatibility: \( \text{Comp}_{NNE} = \{(u, v) \in V \times V : (u, v, -1) \notin E\} \).

DPE is the strictest form of compatibility, while NNE is the most relaxed one. Specifically, \( \text{Comp}_{DPE} \) is the minimal subset of pairs of nodes that satisfies the positive edge compatibility property, while \( \text{Comp}_{NNE} \) is the maximal subset of pairs of nodes that satisfies the negative edge incompatibility property.

We will now use the theory of structural balance [2, 4, 7] to provide more refined definitions of compatibility. The theory is based on the following socially and psychologically founded premises: (1) the friend of my friend is my friend, (2) the friend of my enemy is my enemy, and (3) the enemy of my enemy is my friend. Let \( P = (v_0, \ldots, v_k) \), \((u, v, +1) \in E \) denote a path between nodes \( u \) and \( v \) in a signed graph \( G \). We define the sign of the path as \( \text{sign}(P) = \prod_{i=0}^{k} \text{sign}(v_i, v_{i+1}) \). We say that path \( P \) is positive if \( \text{sign}(P) = +1 \) and negative if \( \text{sign}(P) = -1 \).

**Claim 1.** A positive path \( P_u \) between two nodes \( u \) and \( v \) indicates compatibility, while a negative one indicates incompatibility.

The claim follows from the basic principle of structural balance. To see this, let \( P_u = (x_0, x_1, \ldots, x_k, x_{k+1}) \), \( x_0 = u, x_{k+1} = v \) be a path that connects \( u \) and \( v \). Let \( E_u \) be the set of friends of \( u \) and \( E_v \) be the set of enemies of \( v \). We start by placing node \( u \) in \( F_u \) and traverse the passes as follows. When we traverse edge \((x_i, x_{i+1})\), if the edge is positive we place \( x_{i+1} \) in the same set as \( x_i \). That is, the friends of my friends are also my friends, and the friends of my enemies are my enemies. If the edge is negative, then we place \( x_{i+1} \) in the opposite set of \( x_i \). That is, the enemies of my friends are my enemies, and the enemies of my enemies are my friends. If the path \( P_u \) is positive then \( v \) will be placed in \( E_v \), while if the path is negative it will be placed in \( E_u \).

We first look at shortest paths. We use \( SP_{DPE} \) to denote the set of shortest paths between nodes \( u \) and \( v \). \( SP_{NNE} \) denotes the positive, and \( SP_{NNE} \) the negative ones.

**Definition 3.3.** Shortest Path (SP) compatibility relations: – All Shortest Path (SPA) compatibility: \( \text{Comp}_{SPA} = \{(u, v) \in V \times V : \forall P_{uv} \in SP_{DPE}, \text{sign}(P_{uv}) = +1\} \).

– Majority Shortest Path (SPM) compatibility: \( \text{Comp}_{SPM} = \{(u, v) \in V \times V : |SP^+_{uv}| \geq |SP^-_{uv}|\} \).

– One Shortest Path (SPO) compatibility: \( \text{Comp}_{SPO} = \{(u, v) \in V \times V : \exists P_{uv} \in SP_{DPE}, \text{sign}(P_{uv}) = +1\} \).

We further relax compatibility by asking for positive paths that are not necessarily the shortest ones. Based on structural balance, certain triangles are more stable. A general signed graph is structurally balanced, if it does not contain any cycle with an odd number of negative edges [7].

Given a path \( P \), let \( G_P = (P, E[P]) \) be the graph induced by the nodes of \( P \). We say that path \( P \) is structurally balanced if the subgraph \( G_P \) is structurally balanced. Let \( BP_{DPE} \) denote the set of all structurally balanced paths between \( u \) and \( v \).

### Algorithm 1

The SP-compatibility algorithm.

**Input:** Signed graph \( G \), query node \( q \).

**Output:** The number of positive and negative shortest paths from \( q \) to all other nodes in the graph.

1. Initialize \( N^+(q) = 1 \), \( N^-(q) = 0 \) \( N^+(x) = N^-(x) = 0 \), \( L(q) = 0 \), \( L(x) = \infty \), empty queue \( Q \).
2. \( Q \text{~enque}(q) \).
3. while \( Q \neq \emptyset \) do
   4. \( u = Q\text{-dequeue} \).
   5. for \( x \) adjacent to \( u \) do
   6. if \( L(u) + 1 \leq L(x) \) then
      7. if \( x \notin Q \) then
         8. \( L(x) = L(u) + 1 \).
         9. \( Q\text{-enqueue}(x) \).
      10. if \( \text{sign}(u, x) = +1 \) then
         11. \( N^+(x) = N^+(u); N^-(x) = N^-(u) \).
         12. else if \( \text{sign}(u, x) = -1 \) then
             13. \( N^-(x) = N^+(u); N^+(x) = N^-(u) \).
   14. return \( (N^+, N^-, L) \).
paths having the prefix property. We will use SBP to denote the compatibility relation computed by the exact exhaustive algorithm, and $SBP_\text{H}$ the output of the heuristic algorithm.

4 TEAM FORMATION

We now present algorithms for the TFSN problem. Recall that our goal is to find a team of compatible users, that covers all skills, and minimizes the communication cost. The communication cost is defined as the largest distance between any two pairs of users in the team. We define the distance between two users looking at the positive paths connecting them. Specifically, for DPE and SP compatibility, distance is the length of the shortest path, while for SBP the length is the shortest structurally balanced path. For NNE compatibility, since there be no positive paths, we define distance as the length of the shortest path ignoring its sign.

Algorithm 2 is a generic algorithm that incrementally builds a solution, each time considering an uncovered skill and adding a compatible user having this skill, until all skills are covered. There are two placeholders in this algorithm. The first is the policy for selecting a skill (lines 3 and 8), and the second the policy for selecting a candidate user (line 9).

We consider two policies for selecting skills: select the rarest skill first (as in [9]), and select the least compatible skill first. We define the compatibility degree $cd(s)$ of skill $s$ based on the compatibility between the users with skill $s$ and the users with all other skills: $cd(s) = \sum_{s_j \neq s} g(s, s_j)$, where $g(s, s_j) = \{(u_i, u_j) \in \text{Comp}, s \in \text{skills}(u_i) \land s_j \in \text{skills}(u_j)\}$. most compatible among the remaining users. The first selection aims at minimizing the cost, while the second at maximizing the chances of finding a group of compatible users. We experimentally evaluate different combinations of these policies in Section 5.

5 EXPERIMENTAL EVALUATION

In this section, we compare the different compatibility relations on real datasets and evaluate the team formation algorithm. Datasets. Table 1 details our real-world datasets. Slashdot contains information about users and their posts on Slashdot. We obtained a network of Slashdot users [11], where users have tagged their relationships as friend or foe. Then we used the categories of users’ posts as skills. Epinions contains information about users and their reviews about products. The dataset is created by combining a signed network of Epinions users [11] with the RED dataset which contains information about the products and product categories the users have reviewed. We used the unique user ids to match users in the two datasets, and we assigned as skills to users the categories of the products they have reviewed. Wikipedia [11] is a signed network of editors. The edge sign corresponds to a positive or negative vote in admin elections. Since there was no skill information, we assigned synthetically generated skills to its users. We generated 500 distinct skills with frequencies following a Zipf distribution as in real data. Each skill is assigned to users in the network uniformly at random.

Compatibility Relations. In Table 2, we report the percentage of compatible pairs of users and skills. The DPE is excluded from our analysis, since this corresponds to finding cliques and team formation is too restrictive. Two skills $s_1$ and $s_2$ are compatible if they have compatibility degree $cd(s_1, s_2) > 0$, i.e., there is at least one compatible pair of users $(u, v)$ such that $u$ has $s_1$ and $v$ has $s_2$ (including self-compatibility, if the same user has both skills).

As expected, the number of compatible user and skill pairs increases as we relax the notion of compatibility. For SPA, less than half of the pairs of nodes are compatible, and as low as 21.98% for the case of Wikipedia. Also, in most cases, for a sizeable fraction of pairs of skills there are no compatible users, indicating that for many skill combinations there can be no compatible team. Another interesting observation is that the fraction of compatible pairs for SBP is comparable with that for NNE. This means that, for all pairs that are not directly connected with a negative edge, there exists at least one positive structurally balanced path that connects them.

Table 2: Comparison of compatibility relations

<table>
<thead>
<tr>
<th></th>
<th>SPA</th>
<th>SPM</th>
<th>SPO</th>
<th>SBP</th>
<th>$SBP_\text{H}$</th>
<th>NNE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slashdot</td>
<td>44.72</td>
<td>55.72</td>
<td>72.45</td>
<td>97.85</td>
<td>99.38</td>
<td>99.64</td>
</tr>
<tr>
<td>comp. users</td>
<td>80.57</td>
<td>86.19</td>
<td>92.63</td>
<td>99.11</td>
<td>99.47</td>
<td>99.50</td>
</tr>
<tr>
<td>comp. skills</td>
<td>4.13</td>
<td>4.37</td>
<td>4.57</td>
<td>4.95</td>
<td>4.97</td>
<td>4.53</td>
</tr>
<tr>
<td>average distance</td>
<td>3.84</td>
<td>3.82</td>
<td>3.87</td>
<td>3.97</td>
<td>3.83</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Comparison with unsigned team formation

<table>
<thead>
<tr>
<th></th>
<th>SPA</th>
<th>SPM</th>
<th>SPO</th>
<th>SBP</th>
<th>$SBP_\text{H}$</th>
<th>NNE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ignore sign</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>26%</td>
<td>30%</td>
<td></td>
</tr>
<tr>
<td>Delete negative</td>
<td>0%</td>
<td>2%</td>
<td>18%</td>
<td>66%</td>
<td>76%</td>
<td></td>
</tr>
</tbody>
</table>
6 RELATED WORK

To the best of our knowledge, our work is the first to address team formation in signed networks.

Team Formation. Lappas et al. [9] were the first to formally define the problem of finding a team of experts using the network structure to quantify the quality of the team as a whole. There is considerable amount of work extending their model, (e.g., [1, 8, 12]), but none of these works considers a signed network.

Signed Networks. There is a fair amount of work on signed networks [13]. A problem somehow related to our work is that of link and sign prediction [3, 11]. However, we differentiate, since we are not interested in predicting future links, but rather in evaluating the compatibility between any two individuals in the network. There is also work on detecting communities in signed networks (e.g., see [14]). The notion of the team is somehow related to that of the community, but the objective of team formation is different.

Structural Balance. There is a rich literature in psychology on positive and negative relations among groups of people using structural balance theory, e.g., [2, 4]. Structural balance has been used e.g., for identifying clusters [5] and polarization in networks [10], finding communities [6].

7 CONCLUSIONS

In this paper, we introduced the novel problem of team formation in a signed network. The problem poses the challenge of defining node compatibility in a signed network. To this end, we provided a principled framework by utilizing the theory of structural balance. In the future, we plan to investigate different ways to combine compatibility and communication cost and to exploit compatibility for other tasks, such as link prediction or clustering.

ACKNOWLEDGEMENTS

The research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant” (Project Number: 1873).

REFERENCES


