Mining Weighted Leaders and Peripheral Workers in Organizational Social Networks based on Event Logs

Alessandro Berti

Abstract—Identifying important, influent individuals in a social network has been, for decades, an interesting analysis, that can lead in business contexts to a better understanding of the community structure and workers’ behavior (considering, e.g., performance). In this paper, the focus is on social networks extracted from event logs, and a more powerful definition of leadership is introduced taking into account the fact that leaders may have different importance inside the organization. This concept is useful also in identifying peripheral workers, that are far from leaders. In an assessment done on the BPI Challenge 2012 event log, peripheral workers showed better performance in comparison to other workers. This discovery has been explained using Social Psychology concepts and considering several characterizations of peripheral workers.

Keywords—Weighted Leaders; Peripheral Workers; Clustering; Social Network Analysis; Sociology.

I. INTRODUCTION

An important information about the social structure of an organization regards leadership. A leader is a person that holds a dominant or superior position within its field, and is able to exercise a high degree of control or influence over others. Much emphasis has been given by the literature to the importance of leaders in the functioning of an organization. Research on leader-member exchange [1] shows that normal workers’ performance is influenced by the relation with their leader. Other studies [2][3] show that effective leaders can be generally found in the center of a social network, according to a centrality measure (see for instance [4][5]). These studies are sociometric. Sociometry [6] is a science that can be applied in business contexts, like performance management [7][8]. Blondel et al. [9] is particularly interesting because it speaks about a method to find social leaders inside an organization and how to use that information to increase the business insights (improving the community structure and the graph visualization).

In this paper, the focus has been switched from a sociometric approach (based only on relations between individuals) to an event log approach, because it could be more powerful. In doing so, only social networks extracted from event logs [10] have been considered; they are particularly meaningful in business contexts [11]. The notion of weighted leaders will be explained (in Section 3), a simple algorithm to find them will be introduced (in Section 3) and, moreover, an improved clustering algorithm, inspired by the [9] one, will be presented (in Section 3). The concept of peripheral workers will then be defined (always in Section 3), showing that they usually perform better than other workers (in Section 4), exploring thanks to existing Social Psychology literature [12][13] various types of peripheral workers.

II. BACKGROUND

Social networks in business contexts [10] may be built upon event logs [14][15], which are collections of information about events happening in the organization. These include the event’s timestamp; the process instance in which the event is deployed; the event’s originator (i.e., the worker who does the event). A point that may need to be clarified is that, in Business Process Intelligence [16] terminology (BPI is the analysis of business processes using IT systems), an event is always instantaneous. The concept that many might be familiar with is the one of activity. To understand the difference, To understand the difference, we could think of “Cooking a pasta” as an activity built of possibly two events: a start (instantaneous) event and a completion event (in which, we declare to have already cooked the pasta).

Like in [17], a social network can be defined as a weighted graph \( G = (V, E) \), where nodes represent individuals (workers), and are identified by integers (thus \( V \), the set of nodes, is a subset of \( \mathbb{N} \)); edges represent relations between individuals, and are identified by couples \( e = (i, j) \) (where \( i \) and \( j \) are identifiers of nodes; the set of edges \( E \) is a subset of \( V \times V \)); weights are associated to edges, and are the strength of the relationship represented by the corresponding edge (mathematically, they can be understood as functions from \( E \) to \( \mathbb{R} \)). Given an edge \( (i, j) \in E \) the associated weight is denoted as \( w((i, j)) \in \mathbb{R} \).

To effectively build the social network, a weight (between 0 and 1) to relations between individuals has to be assigned. This can be done calculating a metric between individuals. Van der Aalst et al. in [10] propose several metrics, like the Handover of Work (HoW) metric, that measures how many times the work of an individual for a process instance is followed by the work of another individual; and the Working Together (WT) metric, that measures how many times two individuals work together in process instances. In this paper, the focus will be mainly on the WT metric, as the collaborative distance between leaders and other individuals is considered \((WT(p_1, p_2) = \frac{1}{\text{number of instances in which both } p_1 \text{ and } p_2 \text{ do events and the number of instances, contained in the log, in which } p_1 \text{ do events})\). So, the value of the metric is high when two individuals often collaborate.

Information can be mined from a social network using a clustering algorithm, which groups individuals based on their similarity, to extract information about the community structure of the organization [18][19][20]. A clustering \( C \) of \( G \) is a family of subsets of \( V \) such that each node is assigned to exactly one cluster and a function \( C : V \to \mathbb{N} \) where \( C(v) = i \iff v \in S_i \) \((v \text{ belongs to the cluster } S_i)\) can be
defined. There are several clustering algorithms [19][21]-[26], but unfortunately the majority of them work on undirected graphs. So, to use them on directed graphs, the graph has to be transformed into an undirected one (i.e., making edges (i, j) and (j, i) to have the same weight).

A difficult task is to evaluate the quality of the output of clustering algorithms. In the context of social networks, the most popular criteria to judge the quality of a clustering is modularity. Modularity is a concept, described in [21], that aims to measure group cohesion inside communities and separation between them. The higher the quality, the better the quality of the detected communities is. Some clustering algorithms try to maximize directly modularity (e.g., [21]). Also nodes centrality (degree centrality [27], pagerank centrality [28], betweenness centrality [5]) may be an important factor to understand which individuals are important in their group and to find overlapping communities [19][29].

Having an event log, however, means having more information than the ones contained in Social Networks extracted from the metrics: a Business Process Improvement analysis can be done [15]. An interesting analysis might regards instances completion times. Indeed, instances with an high duration may be dangerous (for example, breaking Service Level Agreement); while the ones with low duration may signal some positive things inside the organization. This concept, in Lean Manufacturing terminology, is called Lead Time [30]. Indeed, focusing on a process, the mean (M) completion time of instances, the standard deviation (SD) of completion times can be calculated, and after fixing a constant k (as example, k = 1) one can consider as “positive” instances the ones whose duration is below M − k ∗ SD, as “normal” instances the ones whose duration is between M − k ∗ SD and M + k ∗ SD, as “negative” instances, or instances whose duration exceeds Lead Time, the ones whose duration is above M + k ∗ SD.

Another interesting Lean Manufacturing-inspired concept is the Flow Rate. It measures the ratio of the quantity of time in which the instance is actively worked and the instance duration. In other words, it is a measure of how many long “holes” there are between the completion of an activity and the start of the next activity. So, instances with lower Flow Rate are being worked in a more systematic way.

III. WEIGHTED LEADERS AND PERIPHERAL WORKERS

In this section, we will define the concepts of weighted leaders and peripheral workers, and we will propose a method to find weighted leaders.

A. Weighted Leaders

Blondel et al. [9] have introduced a method to discover leaders. However, the authors do not consider the fact leaders in an organization may have different weights, i.e., there are leaders which are more important than others. **Definition:** A weighted leader is a couple (i, w) where i is a (leader) individual and w is the weight (comprised between 0 and 1) that measures the importance of the leader.

This is meaningful because less important leaders may have a less wide “sphere of influence” than the most important ones, and this observation can be used to improve the community structure (clustering). Indeed, a clustering algorithm is proposed, inspired to the one described in [9] and reported in Fig. 1, that takes into account weighted leaders. It is described in Fig. 2, and consists in inserting each node in the cluster of its most (weighted) near leader. This method takes into account both the (topological) distance and the power / weight of the leader. In the Assessment section, there is a comparison between this algorithm and the one presented in [9].

B. Peripheral Workers

The proximity of a worker to other workers expresses how much the given worker is profoundly embedded in the organization, and is expressed by the weight of the connections of the given worker to other workers. Having introduced the notion of (weighted) leader, there is interest in observing which workers are far from leaders.

Peripheral workers are workers that are far, in the sense of collaboration, from leaders. They can be found by calculating for each worker a quantity, that is called leader proximity, expressing the distance of the worker from the leaders. The algorithm to calculate leader proximity, and to discover peripheral workers, is described in Fig. 3: the minimum topological distance from a leader, considering also his weight, is found.

The peripheral workers concept is not strictly coincident with other Social Psychology concepts, but two possible categories of peripheral workers can be considered:

- Newcomers are workers that are new in the organization, or were previously assigned to different processes. They can feed new energy to the organization, and new ideas (see [12][31][32]). However, they can be considered marginal in the organization because a new worker usually does not suddenly collaborate with organizational leaders, and his initial collaboration network is usually strict. To enhance their position in the organization, they usually start working harder than their
Weighted_Clustering\((G, L_W)\)

**Require:** A weighted social network graph \(G = (V, E, w)\)

A set of weighted leaders \(L_W = \{(l_1, w_1), \ldots, (l_n, w_n)\}, l_i \in V \forall i\)

**Ensure:** A clustering \(C : V \rightarrow \mathbb{N}\) of \(G\)

\[\forall \pi_1\text{ is the projection on the first component}\]

\[L \leftarrow \pi_1(L_W) \cup L\] is the set of leaders, considered without weight \(L_0\) \(\triangleright\) Clustering, initially empty

\[\text{new}_C \leftarrow \emptyset \triangleright\text{Ausiliar clustering, initially empty}\]

\[W_L \leftarrow L_W \triangleright\text{Leader proximity of workers, initially equal to the weighted leaders set}\]

\[\text{for all } l \in L\ do\]

\[i \leftarrow 0\]

\[\text{for all } i \in L\ do\]

\[n \leftarrow (i + 1)\]

\[\text{new}_C(l) \leftarrow i\]

**end for**

**end for**

while \(\text{new}_C \neq C\) do

\[C \leftarrow \text{new}_C\]

**for all** \(n \in V \setminus \pi_1(C)\) do

\[L_n \leftarrow \{(k, w((n, k))) \mid (n, k) \in E\} \triangleright w(e)\] is the weight of the edge

\(\triangleright\) The following is different from the Blondel’s algorithm

\[\text{new}_C(n) \leftarrow C(l_n)\]

\[W_L' \leftarrow W_L \cup (n, W_L(l_n) \ast w((n, l_n)))\]

\(\triangleright\) In the leader proximity set, the worker with its leader proximity have been inserted

**end for**

**end for**

\(\triangleright\) After that, isolated nodes are inserted

**for all** \(n \in V \setminus \pi_1(C)\) do

\[i \leftarrow (i + 1)\]

\[C(n) \leftarrow i\]

**end for**

**end while**

**return** \(C\)

---

Peripheral Arbeits(G, L_W, t)

**Require:** A weighted social network graph \(G = (V, E, w)\)

A set of weighted leaders \(L_W = \{(l_1, w_k), \ldots, (l_n, w_n)\}, l_i \in V \forall i\)

A threshold \(t\) for peripheral workers

**Ensure:** A set of peripheral workers \(P\)

\(\triangleright\) \(\pi_1\) is the projection on the first component

\[L \leftarrow \pi_1(L_W) \cup L\] is the set of leaders, considered without weight \(L_0\) \(\triangleright\) Peripheral proximity of workers, initially empty

\[P \leftarrow \emptyset \triangleright\text{Peripheral proximity of workers, initially empty}\]

\[\text{new}_P \leftarrow L_W \triangleright\text{Ausiliar set of workers’ leader proximity, initially equal to the set of weighted leaders}\]

while \(\text{new}_P \neq P\) do

\[\text{new}_P \leftarrow P\]

**for all** \(n \in V \setminus \pi_1(L_W)\) do

\[L_n \leftarrow \{(k, w((n, k))) \mid (n, k) \in E\} \triangleright w(e)\] is the weight of the edge

\[l_n \leftarrow \pi_1(\text{arg max}_{E'} \pi_2 \ast W_L(P_1))\]

\[W_L' \leftarrow W_L \cup (n, W_L(l_n) \ast w((n, l_n)))\]

\(\triangleright\) In the leader proximity set, the worker with its leader proximity have been inserted

**end for**

**end while**

**return** \(P\)

---

Figure 2. The algorithm to cluster organizational social networks, having in input the set of weighted leaders

---

Copyright (c) IARIA, 2016. ISBN: 978-1-61208-519-7

---

Figure 3. The algorithm to discover peripheral workers in a social network, having in input the set of weighted leaders and a threshold (for peripheral workers).

---

Blondel Arbeits(G)

**Require:** A weighted social network graph \(G = (V, E, w)\)

**Ensure:** A set of social leaders \(L\)

\(L \leftarrow \emptyset \triangleright\text{Set of leaders, initially empty}\)

**for all** \(n \in V\) do

\[N_n \leftarrow \{k \mid (n, k) \in E\} \setminus \{C(n)\}\] Compute the set of different nodes in the neighborhood of \(n\)

**if** \(N_n \neq \emptyset\) then

\[k_{\text{Leader}} \leftarrow 1\]

**for all** \(k \in \text{Neighborhood}(n)\) do

\(\triangleright\) 3-clique of \(G\) which \(k\) belongs to

\[3-cl(k)\] counts the number of 3-cliques in \(G\) which \(k\) belong to

**if** \(3-cl(k) > 3-cl(n)\) then

\[k_{\text{Leader}} \leftarrow 0\]

**end if**

**end for**

**if** \(k_{\text{Leader}} = 0\) then

\[L \leftarrow L \cup \{n\}\]

**end if**

**end for**

**return** \(L\)

---

Figure 4. Blondel’s algorithm to discover leaders in a social network
Weighted_Leaders(\textit{LOG}, \textit{E})

Require: An event log \textit{LOG}

A weighted social network graph \(G = (V, E, w)\)

Ensure: A set of weighted leaders \(L_{W}\)

\(L_{W} \leftarrow \emptyset \) \(\triangleright \) Set of weighted leaders, initially empty

\(N \leftarrow \emptyset \) \(\triangleright \) Number of instances for worker, initially empty

\(\text{modularities} \leftarrow \emptyset \) \(\triangleright \) Set of modularities for different number of weighted leaders

\(n_{\text{max}} \leftarrow \max_{w \in V} n(w) \triangleright \) The greatest number of instances in \textit{LOG}

in which a single worker collaborated

\textbf{for all} \( w \in V \) \textbf{do}

\(N(w) \leftarrow n(w) \triangleright \) Count the number of instances in \textit{LOG} in which the worker \( w \) does something, \(\triangleright\) and do the ratio with \(n_{\text{max}}\)

\textbf{end for}

order\_decreasing(\(N\))

\textbf{for} \( i = 1, \ldots, |V| \) \textbf{do}

\(L_{\text{temp}} \leftarrow \text{take\_first}(N, i) \triangleright \) Take first \( i \) elements in accordance to ordering

\(\text{modularities} \leftarrow (i, \text{modularity}(\text{Weighted\_Clustering}(G, L_{\text{temp})))) \triangleright \) It computes the modularity of the clustering obtained using the proposed algorithm

\textbf{end for}

\(\triangleright \pi_1\) is the projection on the first component

\(i_{\text{max}} \leftarrow \pi_1(\arg \max_{\text{modularities}} \pi_2)\)

\(L_{W} \leftarrow \text{take\_first}(N, i_{\text{max}})\)

\textbf{return} \(L_{W}\)

Figure 5. The algorithm to discover weighted leaders in a social network.

---

An interesting theory related to this is \textit{job embeddedness} [42]. This theory explains why workers wish to be included in an important network of relations inside the organization, as the ones most embedded in the organizational social network (i.e., having strong ties with other workers, and with leaders) have the best chance to retain the work. Or, finally, they could be members of the group that once were full members but lost their position because they failed to live up with expectations of the group (these workers were studied in [43]). This could explain their good performances as an attempt to being considered again. Basing always on [43], if they succeed in re-doing a socialization, they may resume their activities as full members.

The previous one should not be considered as conclusive categories, but are useful to categorize part of the peripheral workers, while some ones are out of these categories. In the assessment we will see that peripheral workers offer however a good performance, and this could be explained also by other reasons. Peripheral workers, given their marginality in the organization, are assigned to simpler instances. These,
In this section a method is proposed to discover leaders, and to assign them a weight that takes into account the social network and the event log. It is a very simple way, with some approach (including Social Network Analysis).

the analysis. present in almost all instances, and it has been excluded from a metric between individuals: the Working Together metric has been chosen. Worker a worker. The workers with the greater number of worked instances by the worker with the greatest number of worked instances have the ratios between their number of worked instances and the number of process instances.

But how many of the workers, given they have been sorted by that number, should be taken? One should consider the number of the leaders that, according to the clustering algorithm that has been previously introduced, maximizes the quality of the obtained community structure, measured by modularity. Indeed, the number of leaders that realize the maximum represent possibly a good and synthetic covering of the social network graph.

IV. ASSESSMENT

The assessment has been done on the Business Process Intelligence Challenge 2012 event log. This event log, taken from a Dutch financial institute and regarding an application process for personal loans, has been made freely available to invite business process mining specialists to work on discovering possibly interesting business analysis, using any available approach (including Social Network Analysis).

As explained in the background, social networks extracted from event logs are being considered, so it is required to choose a metric between individuals: the Working Together metric has been chosen. Worker 112 is an automated resource, that is present in almost all instances, and it has been excluded from the analysis.

Table I resumes the obtained results, for the considered

In Fig. 4) is briefly recalled: given a node (worker), if the number of 3-cliques (a N-clique is a subset of size N of the vertices such that every two distinct vertices are adjacent) it belongs to exceeds the number of 3-cliques neighbors (workers) nodes (workers) belong to, then it is considered to be a social leader.

The approach proposed in this paper (described in Fig. 5) is focused on counting the number of process instances worked by the workers. The resources with the greater number of process instances are considered to be leaders. The weight is 1 for the worker between the greatest number of process instances, and for other workers that are considered to be leaders, is the ratio between their number of worked instances and the number of worked instances by the worker with the greatest number of process instances.

But how many of the workers, given they have been sorted by that number, should be taken? One should consider the number of the leaders that, according to the clustering algorithm that has been previously introduced, maximizes the quality of the obtained community structure, measured by modularity. Indeed, the number of leaders that realize the maximum represent possibly a good and synthetic covering of the social network graph.

IV. ASSESSMENT

The assessment has been done on the Business Process Intelligence Challenge 2012 event log. This event log, taken from a Dutch financial institute and regarding an application process for personal loans, has been made freely available to invite business process mining specialists to work on discovering possibly interesting business analysis, using any available approach (including Social Network Analysis).

As explained in the background, social networks extracted from event logs are being considered, so it is required to choose a metric between individuals: the Working Together metric has been chosen. Worker 112 is an automated resource, that is present in almost all instances, and it has been excluded from the analysis.

Table I resumes the obtained results, for the considered

In Fig. 4) is briefly recalled: given a node (worker), if the number of 3-cliques (a N-clique is a subset of size N of the vertices such that every two distinct vertices are adjacent) it belongs to exceeds the number of 3-cliques neighbors (workers) nodes (workers) belong to, then it is considered to be a social leader.

The approach proposed in this paper (described in Fig. 5) is focused on counting the number of process instances worked by the workers. The resources with the greater number of process instances are considered to be leaders. The weight is 1 for the worker between the greatest number of process instances, and for other workers that are considered to be leaders, is the ratio between their number of worked instances and the number of worked instances by the worker with the greatest number of process instances.

But how many of the workers, given they have been sorted by that number, should be taken? One should consider the number of the leaders that, according to the clustering algorithm that has been previously introduced, maximizes the quality of the obtained community structure, measured by modularity. Indeed, the number of leaders that realize the maximum represent possibly a good and synthetic covering of the social network graph.

IV. ASSESSMENT

The assessment has been done on the Business Process Intelligence Challenge 2012 event log. This event log, taken from a Dutch financial institute and regarding an application process for personal loans, has been made freely available to invite business process mining specialists to work on discovering possibly interesting business analysis, using any available approach (including Social Network Analysis).

As explained in the background, social networks extracted from event logs are being considered, so it is required to choose a metric between individuals: the Working Together metric has been chosen. Worker 112 is an automated resource, that is present in almost all instances, and it has been excluded from the analysis.

Table I resumes the obtained results, for the considered

In Fig. 4) is briefly recalled: given a node (worker), if the number of 3-cliques (a N-clique is a subset of size N of the vertices such that every two distinct vertices are adjacent) it belongs to exceeds the number of 3-cliques neighbors (workers) nodes (workers) belong to, then it is considered to be a social leader.

The approach proposed in this paper (described in Fig. 5) is focused on counting the number of process instances worked by the workers. The resources with the greater number of process instances are considered to be leaders. The weight is 1 for the worker between the greatest number of process instances, and for other workers that are considered to be leaders, is the ratio between their number of worked instances and the number of worked instances by the worker with the greatest number of process instances.

But how many of the workers, given they have been sorted by that number, should be taken? One should consider the number of the leaders that, according to the clustering algorithm that has been previously introduced, maximizes the quality of the obtained community structure, measured by modularity. Indeed, the number of leaders that realize the maximum represent possibly a good and synthetic covering of the social network graph.

IV. ASSESSMENT

The assessment has been done on the Business Process Intelligence Challenge 2012 event log. This event log, taken from a Dutch financial institute and regarding an application process for personal loans, has been made freely available to invite business process mining specialists to work on discovering possibly interesting business analysis, using any available approach (including Social Network Analysis).

As explained in the background, social networks extracted from event logs are being considered, so it is required to choose a metric between individuals: the Working Together metric has been chosen. Worker 112 is an automated resource, that is present in almost all instances, and it has been excluded from the analysis.

Table I resumes the obtained results, for the considered

In Fig. 4) is briefly recalled: given a node (worker), if the number of 3-cliques (a N-clique is a subset of size N of the vertices such that every two distinct vertices are adjacent) it belongs to exceeds the number of 3-cliques neighbors (workers) nodes (workers) belong to, then it is considered to be a social leader.

The approach proposed in this paper (described in Fig. 5) is focused on counting the number of process instances worked by the workers. The resources with the greater number of process instances are considered to be leaders. The weight is 1 for the worker between the greatest number of process instances, and for other workers that are considered to be leaders, is the ratio between their number of worked instances and the number of worked instances by the worker with the greatest number of process instances.

But how many of the workers, given they have been sorted by that number, should be taken? One should consider the number of the leaders that, according to the clustering algorithm that has been previously introduced, maximizes the quality of the obtained community structure, measured by modularity. Indeed, the number of leaders that realize the maximum represent possibly a good and synthetic covering of the social network graph.

IV. ASSESSMENT

The assessment has been done on the Business Process Intelligence Challenge 2012 event log. This event log, taken from a Dutch financial institute and regarding an application process for personal loans, has been made freely available to invite business process mining specialists to work on discovering possibly interesting business analysis, using any available approach (including Social Network Analysis).

As explained in the background, social networks extracted from event logs are being considered, so it is required to choose a metric between individuals: the Working Together metric has been chosen. Worker 112 is an automated resource, that is present in almost all instances, and it has been excluded from the analysis.

Table I resumes the obtained results, for the considered

In Fig. 4) is briefly recalled: given a node (worker), if the number of 3-cliques (a N-clique is a subset of size N of the vertices such that every two distinct vertices are adjacent) it belongs to exceeds the number of 3-cliques neighbors (workers) nodes (workers) belong to, then it is considered to be a social leader.

The approach proposed in this paper (described in Fig. 5) is focused on counting the number of process instances worked by the workers. The resources with the greater number of process instances are considered to be leaders. The weight is 1 for the worker between the greatest number of process instances, and for other workers that are considered to be leaders, is the ratio between their number of worked instances and the number of worked instances by the worker with the greatest number of process instances.

But how many of the workers, given they have been sorted by that number, should be taken? One should consider the number of the leaders that, according to the clustering algorithm that has been previously introduced, maximizes the quality of the obtained community structure, measured by modularity. Indeed, the number of leaders that realize the maximum represent possibly a good and synthetic covering of the social network graph.

IV. ASSESSMENT

The assessment has been done on the Business Process Intelligence Challenge 2012 event log. This event log, taken from a Dutch financial institute and regarding an application process for personal loans, has been made freely available to invite business process mining specialists to work on discovering possibly interesting business analysis, using any available approach (including Social Network Analysis).

As explained in the background, social networks extracted from event logs are being considered, so it is required to choose a metric between individuals: the Working Together metric has been chosen. Worker 112 is an automated resource, that is present in almost all instances, and it has been excluded from the analysis.

Table I resumes the obtained results, for the considered
<table>
<thead>
<tr>
<th>Worker</th>
<th>Leader proximity</th>
<th>Max worker proximity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10861</td>
<td>1.000</td>
<td>0.308</td>
</tr>
<tr>
<td>11181</td>
<td>0.972</td>
<td>0.327</td>
</tr>
<tr>
<td>11169</td>
<td>0.960</td>
<td>0.278</td>
</tr>
<tr>
<td>10913</td>
<td>0.932</td>
<td>0.308</td>
</tr>
<tr>
<td>10821</td>
<td>0.931</td>
<td>1.000</td>
</tr>
<tr>
<td>11119</td>
<td>0.931</td>
<td>0.300</td>
</tr>
<tr>
<td>11180</td>
<td>0.884</td>
<td>0.359</td>
</tr>
<tr>
<td>11254</td>
<td>0.884</td>
<td>1.000</td>
</tr>
<tr>
<td>10909</td>
<td>0.871</td>
<td>0.305</td>
</tr>
<tr>
<td>11203</td>
<td>0.855</td>
<td>0.315</td>
</tr>
<tr>
<td>11189</td>
<td>0.797</td>
<td>0.298</td>
</tr>
<tr>
<td>11201</td>
<td>0.792</td>
<td>0.367</td>
</tr>
<tr>
<td>10982</td>
<td>0.753</td>
<td>0.272</td>
</tr>
<tr>
<td>10125</td>
<td>0.697</td>
<td>1.000</td>
</tr>
<tr>
<td>11049</td>
<td>0.697</td>
<td>0.313</td>
</tr>
<tr>
<td>11259</td>
<td>0.613</td>
<td>0.314</td>
</tr>
<tr>
<td>11122</td>
<td>0.594</td>
<td>0.271</td>
</tr>
<tr>
<td>10899</td>
<td>0.578</td>
<td>0.302</td>
</tr>
<tr>
<td>11079</td>
<td>0.563</td>
<td>0.563</td>
</tr>
<tr>
<td>10124</td>
<td>0.513</td>
<td>0.600</td>
</tr>
<tr>
<td>11319</td>
<td>0.415</td>
<td>0.485</td>
</tr>
<tr>
<td>10914</td>
<td>0.393</td>
<td>0.444</td>
</tr>
<tr>
<td>11002</td>
<td>0.369</td>
<td>0.396</td>
</tr>
<tr>
<td>11299</td>
<td>0.366</td>
<td>0.428</td>
</tr>
<tr>
<td>11339</td>
<td>0.356</td>
<td>0.615</td>
</tr>
<tr>
<td>10939</td>
<td>0.345</td>
<td>0.345</td>
</tr>
<tr>
<td>10935</td>
<td>0.345</td>
<td>0.370</td>
</tr>
<tr>
<td>11000</td>
<td>0.338</td>
<td>0.348</td>
</tr>
<tr>
<td>11111</td>
<td>0.338</td>
<td>0.348</td>
</tr>
<tr>
<td>11269</td>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>10863</td>
<td>0.328</td>
<td>0.328</td>
</tr>
<tr>
<td>11121</td>
<td>0.324</td>
<td>0.324</td>
</tr>
<tr>
<td>10809</td>
<td>0.324</td>
<td>0.410</td>
</tr>
<tr>
<td>11302</td>
<td>0.323</td>
<td>0.378</td>
</tr>
<tr>
<td>11309</td>
<td>0.314</td>
<td>0.367</td>
</tr>
<tr>
<td>10972</td>
<td>0.313</td>
<td>0.449</td>
</tr>
<tr>
<td>10931</td>
<td>0.313</td>
<td>0.322</td>
</tr>
<tr>
<td>10933</td>
<td>0.310</td>
<td>0.363</td>
</tr>
<tr>
<td>11179</td>
<td>0.308</td>
<td>0.311</td>
</tr>
<tr>
<td>10929</td>
<td>0.302</td>
<td>0.311</td>
</tr>
<tr>
<td>11029</td>
<td>0.297</td>
<td>0.334</td>
</tr>
<tr>
<td>11019</td>
<td>0.289</td>
<td>0.310</td>
</tr>
<tr>
<td>10932</td>
<td>0.287</td>
<td>0.292</td>
</tr>
<tr>
<td>11003</td>
<td>0.285</td>
<td>0.301</td>
</tr>
<tr>
<td>10609</td>
<td>0.282</td>
<td>0.377</td>
</tr>
<tr>
<td>10881</td>
<td>0.281</td>
<td>0.289</td>
</tr>
<tr>
<td>11009</td>
<td>0.279</td>
<td>0.279</td>
</tr>
<tr>
<td>10228</td>
<td>0.278</td>
<td>0.286</td>
</tr>
<tr>
<td>11289</td>
<td>0.274</td>
<td>0.433</td>
</tr>
<tr>
<td>11202</td>
<td>0.274</td>
<td>0.282</td>
</tr>
<tr>
<td>10138</td>
<td>0.266</td>
<td>0.382</td>
</tr>
<tr>
<td>10910</td>
<td>0.265</td>
<td>0.265</td>
</tr>
<tr>
<td>10889</td>
<td>0.261</td>
<td>0.280</td>
</tr>
<tr>
<td>10789</td>
<td>0.261</td>
<td>0.377</td>
</tr>
<tr>
<td>10629</td>
<td>0.255</td>
<td>0.366</td>
</tr>
<tr>
<td>11001</td>
<td>0.248</td>
<td>0.267</td>
</tr>
<tr>
<td>10912</td>
<td>0.245</td>
<td>0.276</td>
</tr>
<tr>
<td>11300</td>
<td>0.244</td>
<td>0.274</td>
</tr>
<tr>
<td>11200</td>
<td>0.234</td>
<td>0.240</td>
</tr>
<tr>
<td>10880</td>
<td>0.228</td>
<td>0.235</td>
</tr>
<tr>
<td>10862</td>
<td>0.221</td>
<td>0.230</td>
</tr>
<tr>
<td>10859</td>
<td>0.219</td>
<td>0.228</td>
</tr>
<tr>
<td>10779</td>
<td>0.215</td>
<td>0.231</td>
</tr>
<tr>
<td>11304</td>
<td>0.194</td>
<td>1.000</td>
</tr>
<tr>
<td>10971</td>
<td>0.192</td>
<td>0.192</td>
</tr>
<tr>
<td>11120</td>
<td>0.156</td>
<td>0.177</td>
</tr>
<tr>
<td>10188</td>
<td>0.098</td>
<td>0.126</td>
</tr>
</tbody>
</table>

Social network, using the proposed algorithms. Leaders (reported in bold) are defined using the criterion explained in Fig. 5, considering the workers having greater leader weight (this was, for completeness, reported for each worker). Then, for each worker, (worked) process instances whose duration exceeded Lead Time (which is set to be \( M + k \cdot SD \), with \( k = 1.5 \)) have been considered, calculating the mean Flow Rate, and reporting also the percentage of "critical" process instances over the number of overall worked instances. Peripheral workers (which are emphasized in italic) are the ones with lower leader proximity (in this table, the ones with measure < 0.24 are considered).

The list of leaders was found using the algorithm described in Fig. 5, with their number established trying to maximize the modularity. For \( N = 15 \) there is a value of modularity equal to 0.02431, which is better than the value of modularity obtained applying the algorithm described in [9] (that produces a modularity of −0.02834). So, the weighted leaders-based algorithm manages to get a better description of the community structure than the algorithm described in [9]. Also, this is not due to the chosen number of leaders: Fig. 6 (line coloured light gray represents modularity results for Blondel’s Algorithm on these differently-sized lists of leaders; line with dark gray colour represents modularity results for our algorithm) shows us that, for any chosen number of leaders, the proposed algorithm works better. It provides, in addition, better modularity results than Label Propagation algorithm (that gets a 0.0000) and Multilevel algorithm (that gets a −0.0186), which are commonly used algorithms.

Using the set of (weighted) leaders, leader proximity has been calculated for all remaining workers. The considered peripheral workers are the ones with leader proximity < 0.24: this threshold was chosen for the log because it separates the ones which are peripheral workers from the other workers in Table II. However, in a different log from BPI Challenge 2012 the ideal threshold for peripheral workers is likely to be different. Peripheral workers have definitely a lower percentage of process instances exceeding Lead Time and lower mean Flow Rate.

In Fig. 7 and 8, peripheral workers are shown to perform better than other workers when the focus is on process instances whose duration exceeds Lead Time (that is set to be \( M + k \cdot SD \), with \( k = 1.5 \)) there is a lower percentage of these instances and the mean Flow Rate is inferior. This is not dependant on the threshold chosen to decide peripheral workers, as in Fig. 7 and 8.

In Table III is shown that in many times peripheral workers are also far from other workers, not only from the leaders. This confirms their substantial marginality in the organization. These workers are not clearly part of any work group in the considered organization (an hypotetisus may be that they are external collaborators), running the risk of social exclusion inside the workplace. Some insights could be given on a possible "classification" of some of the peripheral workers:

Worker 11304 enters the event log very late (Sat Feb 04 2012): this says that, relatively to the given process, he is a newcomer, and he may perform great to let the others know him.

Workers 10859, 11120 and 10880 are present only at the start of the event log: this says that they are not fully part of the organization. A hypothesis could be that they are external collaborators, so they perform well to being "called again" by
In this paper insights on organizational social networks extracted from event logs have been proposed, introducing the concept of weighted leader and showing how to find peripheral workers. Basing on these concepts, a clustering algorithm has been introduced, that is similar to the one described in [9], but is based on the concept of weighted leader. Experiments have been carried out on the freely available BPI Challenge 2012 event log, and the leaders were found based on the simple, but effective, count of the worked process instances. In this log, the proposed clustering algorithm produced the best modularity results, and peripheral workers were found to have better behaviour (relative to Flow Rate and to the percentage of cases exceeding Lead Time) than other workers.

The classification of at least some of the peripheral workers has been proposed, in some categories that come from Social Psychology literature. The analysed categories are the newcomers and the social excluded workers. In the log there is at least one peripheral worker for both the categories, and this lead to possible explanations about the good behaviour (relatively to the considered performance measures) of these workers.

For an organization, given newcomers good behaviour, it may be convenient to involve new people in (existing) processes, rather than being fixed on a static workforce [31]. Also, recurring to external collaborators and contracts that expires seems convenient in order to avoid workers to be static on a process and, in the long period, to lose motivation and performance [34][35][36].

The practical purpose of the proposed methods is to easily find, starting from an event log, high and low performing workers, doing an effective evaluation of employees. The proposed algorithms work on event logs: only few organizations have a process-awareness level such that they collect data in an event log, through Information Technology systems. These logs often are private: there is a very little number of freely available event logs and the chosen one (BPI Challenge 2012, that collects event from a Dutch financial institute) is probably one of the most meaningful.

Also, only events regarding a particular process (application process for personal loans) have been inserted in BPI Challenge 2012: this limits the social network to the people working for that particular scope. In addition to that, one does not know anything other than the information written in the event log. This has lead to a classification of peripheral workers that is plausible but must be seen like an hypothesis, that could not be confirmed given the information in the BPI Challenge 2012 log.

It must also be remarked that information obtained here regard only a process. So, the found leaders and peripheral workers might be leaders, or marginals, only in the given process, not necessarily in the organization.

An open question regards the possibility of a better criterion to discover leaders in the social network. Some ideas that are based on possible calculations on the event log are about introducing some measures, and they are briefly reported:

- A statistical measure of workload (number of things done contemporaneously), searching leaders among the workers having greater workload.
- A notion of criticality among workers, that is high when a worker does a type of activity with no or few possible replacements in the organization. Leaders often do exclusive activities, because of their role, so a good criterion to discover leaders may be calculating workers’ criticality and taking the ones with the higher measure.
- Measuring responsibility through the in-degree of the worker in the Handover of Work between-individuals metric [10]. Indeed, if there are many handovers, it means that many workers in the organization need to consult the given individual, so there is a good possibility that he is a leader.
- Measuring worktime. Leaders usually have more responsibilities, so they work longer.

Analyzing the effect of these ideas, however, is a big task, that goes beyond the purposes of this paper and, given the effectiveness of the measure that counts the number of cases, is left as future work.

REFERENCES