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

Alessandro Berti

SIAV
35030 Rubano PD
Email: alessandro.berti89@gmail.com

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) is the ratio of the number of instances in which both p_1 and p_2 do events and the number of instances, contained in the log, in which p_1 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 belongs to the cluster S_i) can be

 ISBN: 978-1-61208-519-7

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 is the modularity, 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 Agreements); 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 $M - k \cdot SD$, as "normal" instances the ones whose duration is between $M - k \cdot SD$ and $M + k \cdot SD$, as "negative" instances, or instances whose duration exceeds Lead Time, the ones whose duration is above $M + k \cdot SD$.

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].

Blondel_Clustering(G, L)
Require: A weighted social network graph $G = (V, E, w)$
  A set of leaders $L = \{l_1, \ldots, l_n\}$, $l_i \in V\ \forall i$
Ensure: A clustering $C : V \to \mathbb{N}$ of G
  $C \leftarrow \emptyset$ ▷ Clustering, initially empty
  new_C $\leftarrow \emptyset$ ▷ Ausiliar clustering, initially empty
  $i \leftarrow 0$
  for all $l \in L$ **do**
    $i \leftarrow (i + 1)$
    new_C$(l) \leftarrow i$
  end for
  while *new_C* $\neq C$ **do**
    $C \leftarrow$ *neq_C*
    ▷ π_1 is the projection on the first component
    ▷ So, roughly speaking, I are taking the nodes
    for all $n \in V \setminus \pi_1(C)$ **do**
      $L_n \leftarrow \{(k, w((n, k))) \mid (n, k) \in E\}$ ▷ $w(e)$ is the weight of the edge
      $l_n \leftarrow \pi_1(\arg\max_{L_n} \pi_2)$
      new_C$(n) \leftarrow C(l_n)$
    end for
  end while
  ▷ After that, isolated nodes are inserted
  for all $n \in V \setminus \pi_1(C)$ **do**
    $i \leftarrow (i + 1)$
    $C(n) = i$
  end for
  return C

Figure 1. Blondel's algorithm to cluster organizational social networks, having in input the set of leaders

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

 ISBN: 978-1-61208-519-7

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 $C : V \to \mathbb{N}$ of G
▷ π_1 is the projection on the first component
$L \leftarrow \pi_1(L_W)$ ▷ L is the set of leaders, considered without weight
$C \leftarrow \emptyset$ ▷ Clustering, initially empty
new_C $\leftarrow \emptyset$ ▷ Ausiliar clustering, initially empty
$W_L \leftarrow L_W$ ▷ Leader proximity of workers, initially equal to the weighted leaders set
$i \leftarrow 0$
for all $l \in L$ **do**
  $i \leftarrow (i + 1)$
  new_C$(l) \leftarrow i$
end for
while *new_C* $\neq C$ **do**
  $C \leftarrow$ *neq_C*
  for all $n \in V \setminus \pi_1(C)$ **do**
    $L_n \leftarrow \{(k, w((n, k))) \mid (n, k) \in E\}$ ▷ $w(e)$ is the weight of the edge
    ▷ The following is different from the Blondel's algorithm
    $\mathbf{l_n} \leftarrow \boldsymbol{\pi_1}(\arg\max_{\mathbf{L_n}} \boldsymbol{\pi_2} * \mathbf{W_L}(\boldsymbol{\pi_1}))$
    new_C$(n) \leftarrow C(l_n)$
    $W_L \leftarrow W_L \cup (n, W_L(l_n) * w((n, l_n)))$
    ▷ In the leader proximity set, the worker with its leader proximity have been inserted
  end for
end while
▷ 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
return C

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

mates [33], and this suggests that peripheral workers of this category may offer better performance than other workers. Also, they may motivate old-timers to reflect on the group's work practices [34][35][36], and be a source of diversity regarding the skills and values, which can stimulate the group to consider new ideas and adopt new practices [32][37][38], and this can also contribute to better performances.

- Workers suffering phenomenons similar to *social exclusion* [39], so they are not, or are not considered by other workers, full part of the organizational processes and work force. Social exclusion usually leads to offering a lower performance level [40]. However, possible reasons could be asserted on why peripheral workers may not be full part of the organizational work force, yet offering a good performance level: they are external collaborators or consultants (so they do not always work for the organization). They might have a good working behaviour in order to convince the organization to collaborate again with them. This category contains also workers with expiring contracts that wish to be called again by the organization. A second reason is that they might not feel adequately considered by colleagues, and work hard in order to improve their position in the organization (see [41]).

Peripheral_Workers(G, L_W, t)
Require: A weighted social network graph $G = (V, E, w)$
A set of weighted leaders $L_W = \{(l_1, w_1), \ldots, (l_n, w_n)\}$, $l_i \in V\ \forall i$
A threshold t for peripheral workers
Ensure: A set of peripheral workers P
▷ π_1 is the projection on the first component
$L \leftarrow \pi_1(L_W)$ ▷ L is the set of leaders, considered without weight
$P \leftarrow \emptyset$ ▷ P is the set of peripheral workers, initially empty
$W_L \leftarrow \emptyset$ ▷ Leader proximity of workers, initially empty
new_W$_L \leftarrow L_W$ ▷ Ausiliar set of workers' leader proximity, initially equal to the set of weighted leaders
while *new_W*$_L \neq W_L$ **do**
  $W_L \leftarrow$ *new_W*$_L$
  for all $n \in V \setminus \pi_1(W_L)$ **do**
    $L_n \leftarrow \{(k, w((n, k))) \mid (n, k) \in E\}$ ▷ $w(e)$ is the weight of the edge
    $l_n \leftarrow \pi_1(\arg\max_{L_n} \pi_2 * W_L(\pi_1))$
    $W_L \leftarrow W_L \cup (n, W_L(l_n) * w((n, l_n)))$
    ▷ In the leader proximity set, the worker with its leader proximity is inserted
  end for
end while
▷ After that, isolated nodes are inserted
for all $n \in V \setminus \pi_1(W_L)$ **do**
  $W_L \leftarrow W_L \cup (n, 0)$
end for
for all $(w, v) \in W_L$ **do**
  ▷ v is the leader proximity of worker w
  if $v < T$ **then**
    $P \leftarrow P \cup \{w\}$
  end if
end for
return P

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_Leaders(G)
Require: A weighted social network graph $G = (V, E, w)$
Ensure: A set of social leaders L
$L \leftarrow \emptyset$ ▷ 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 neighbourhood of n
  if $N_n \neq \emptyset$ **then**
    Is_Leader $\leftarrow 1$
    for all $k \in$ *Neigborhood*(n) **do**
      ▷ 3-$cl(k)$ counts the number of 3-cliques in G which k belong to
      if 3-$cl(k) >$ 3-$cl(n)$ **then**
        Is_Leader $\leftarrow 0$
      end if
    end for
    if *Is_Leader* $= 1$ **then**
      $L \leftarrow L \cup \{n\}$
    end if
  end if
end for
return L

Figure 4. Blondel's algorithm to discover leaders in a social network

Weighted_Leaders($\boldsymbol{LOG}, E$)
Require: An event log LOG
A weighted social network graph $G = (V, E, w)$
Ensure: A set of weighted leaders L_W
$L_W \leftarrow \emptyset$ ▷ Set of weighted leaders, initially empty
$N \leftarrow \emptyset$ ▷ Number of instances for worker, initially empty
modularities $\leftarrow \emptyset$ ▷ Set of modularities for different number of weighted leaders
$n_{max} \leftarrow \max_w n(w)$ ▷ The greatest number of instances in LOG in which a single worker collaborated
for all $w \in V$ **do**
$N(w) \leftarrow n(w)$ ▷ Count the number of instances in LOG in which the worker w does something,
▷ and do the ratio with n_{max}
end for
order_decreasing(N)
for $i = 1, \ldots, \mid V \mid$ **do**
$L_{temp} \leftarrow$ *take_first*(N, i) ▷ Take first i elements in accordance to ordering
modularities $\leftarrow$
$(i,$ *modularity*$($*Weighted_Clustering*$(G, L_{temp})))$
▷ It computes the modularity of the clustering obtained using the proposed algorithm
end for
▷ π_1 is the projection on the first component
$i_{max} \leftarrow \pi_1\,(\arg\max_{modularities} \pi_2)$
$L_W \leftarrow$ *take_first*(N, i_{max})
return L_W

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

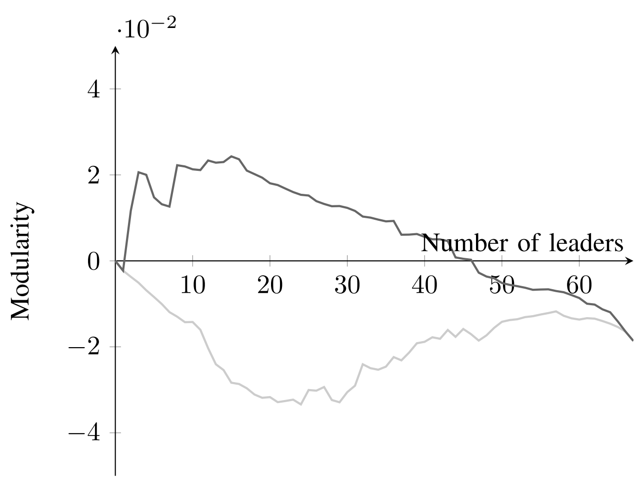

Figure 6. Modularity results of BPI Challenge 2012's Working Together based social network, using Blondel's Leaders-based clustering algorithm and the (weighted) Leaders-based clustering algorithm.

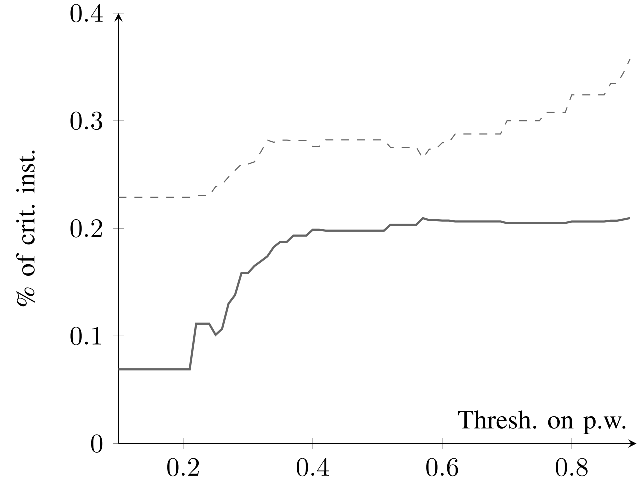

Figure 7. Mean percentage of process instances exceeding Lead Time (which is set to be $M+k{\cdot}SD$ with $k = 1.5$) for peripheral workers and other workers.

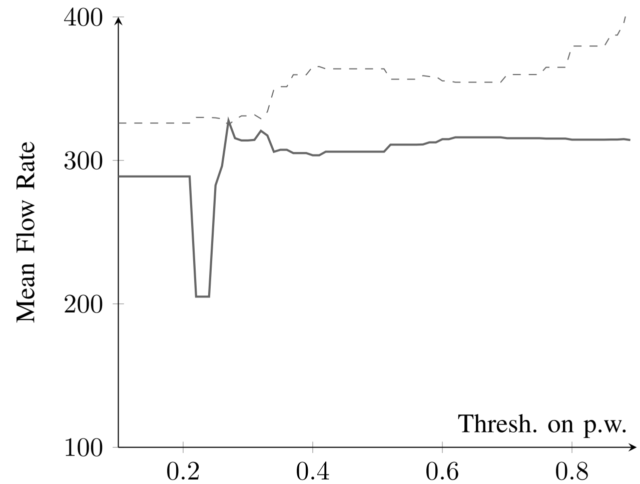

Figure 8. Mean Flow Rate of process instances exceeding Lead Time (which is set to be $M+k{\cdot}SD$ with $k = 1.5$) for peripheral workers and other workers.

An interesting theory related to this is *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,

TABLE I. LIST OF WORKERS IN THE BPI CHALLENGE 2012 EVENT LOG.

Worker	Number of cases worked	Leader weight	Mean Flow Rate	% of critical cases worked	Leader proximity
10861	**1786**	**1.000**	**280.696**	**23.124** %	**1.000**
11181	**1736**	**0.972**	**343.687**	**24.654** %	**0.972**
11169	**1714**	**0.960**	**275.473**	**17.970** %	**0.960**
10913	**1664**	**0.932**	**315.914**	**24.700** %	**0.932**
11119	**1662**	**0.931**	**317.680**	**24.007** %	**0.931**
11180	**1578**	**0.884**	**284.269**	**26.869** %	**0.884**
10909	**1555**	**0.871**	**328.306**	**26.238** %	**0.871**
11203	**1527**	**0.855**	**318.556**	**24.100** %	**0.855**
11189	**1424**	**0.797**	**301.353**	**21.348** %	**0.797**
11201	**1414**	**0.792**	**295.160**	**25.672** %	**0.792**
10982	**1344**	**0.753**	**304.367**	**21.280** %	**0.753**
11049	**1245**	**0.697**	**290.165**	**14.056** %	**0.697**
11259	**1094**	**0.613**	**367.576**	**17.367** %	**0.613**
11122	**1061**	**0.594**	**401.087**	**19.039** %	**0.594**
10899	**1033**	**0.578**	**368.085**	**13.843** %	**0.578**
10881	1026	0.574	399.833	25.536 %	0.281
10138	1022	0.572	365.311	11.057 %	0.266
11179	1003	0.562	341.545	24.128 %	0.308
10932	1000	0.560	328.229	21.400 %	0.287
10910	986	0.552	405.910	15.822 %	0.265
11121	942	0.527	286.044	20.170 %	0.324
11000	914	0.512	251.099	27.790 %	0.338
10609	892	0.499	285.194	14.574 %	0.282
11003	861	0.482	331.269	23.229 %	0.285
10889	786	0.440	312.309	21.883 %	0.261
10972	771	0.432	327.024	13.619 %	0.313
10863	746	0.418	247.067	24.799 %	0.328
10809	744	0.417	303.471	20.699 %	0.324
11009	725	0.406	336.090	20.552 %	0.279
10929	675	0.378	294.035	20.741 %	0.302
10939	647	0.362	315.541	19.784 %	0.345
10629	640	0.358	363.460	13.438 %	0.255
11019	568	0.318	204.465	21.127 %	0.289
10912	536	0.300	280.276	15.485 %	0.245
11202	482	0.270	0.000	0.000 %	0.274
11002	467	0.261	214.214	29.550 %	0.369
10933	405	0.227	395.922	20.494 %	0.310
10789	369	0.207	416.841	17.344 %	0.261
10931	367	0.205	234.828	26.703 %	0.313
11029	365	0.204	0.000	0.000 %	0.297
11200	*341*	*0.191*	*0.000*	*0.000 %*	*0.234*
11120	*294*	*0.165*	*0.000*	*0.000 %*	*0.156*
11289	282	0.158	324.069	17.376 %	0.274
11299	278	0.156	322.202	27.698 %	0.366
10935	265	0.148	340.945	31.698 %	0.345
11300	263	0.147	649.293	6.084 %	0.244
11302	262	0.147	354.620	14.885 %	0.323
11309	229	0.128	451.385	15.721 %	0.314
10880	*226*	*0.127*	*0.000*	*0.000 %*	*0.228*
11319	204	0.114	393.516	16.667 %	0.415
10228	175	0.098	163.990	11.429 %	0.278
10862	*160*	*0.090*	*0.000*	*0.000 %*	*0.221*
10859	*136*	*0.076*	*0.000*	*0.000 %*	*0.219*
10914	135	0.076	251.794	38.519 %	0.393
10971	*130*	*0.073*	*0.000*	*0.000 %*	*0.192*
10188	*87*	*0.049*	*288.814*	*6.897 %*	*0.098*
11001	60	0.034	73.564	6.667 %	0.248
10779	*26*	*0.015*	*121.232*	*15.385 %*	*0.215*
11111	23	0.013	0.000	0.000 %	0.338
11079	16	0.009	316.709	43.750 %	0.563
11339	13	0.007	0.000	0.000 %	0.356
11304	*10*	*0.006*	*0.000*	*0.000 %*	*0.194*
10124	5	0.003	487.185	40.000 %	0.513
11269	3	0.002	42.793	33.333 %	0.333
10125	2	0.001	0.000	0.000 %	0.697
11254	2	0.001	0.000	0.000 %	0.884
10821	1	0.001	952.603	100.000 %	0.931

TABLE II. LIST OF WORKERS IN THE BPI CHALLENGE 2012 EVENT LOG, SORTED INCREASINGLY BY THEIR LEADER PROXIMITY.

Worker	Number of cases worked	Leader weight	Mean Flow Rate	% of critical cases worked	Leader proximity
10188	*87*	*0.049*	*288.814*	*6.897 %*	*0.098*
11120	*294*	*0.165*	*0.000*	*0.000 %*	*0.156*
10971	*130*	*0.073*	*0.000*	*0.000 %*	*0.192*
11304	*10*	*0.006*	*0.000*	*0.000 %*	*0.194*
10779	*26*	*0.015*	*121.232*	*15.385 %*	*0.215*
10859	*136*	*0.076*	*0.000*	*0.000 %*	*0.219*
10862	*160*	*0.090*	*0.000*	*0.000 %*	*0.221*
10880	*226*	*0.127*	*0.000*	*0.000 %*	*0.228*
11200	*341*	*0.191*	*0.000*	*0.000 %*	*0.234*
11300	263	0.147	649.293	6.084 %	0.244
10912	536	0.300	280.276	15.485 %	0.245
11001	60	0.034	73.564	6.667 %	0.248
10629	640	0.358	363.460	13.438 %	0.255
10789	369	0.207	416.841	17.344 %	0.261
10889	786	0.440	312.309	21.883 %	0.261
10910	986	0.552	405.910	15.822 %	0.265
10138	1022	0.572	365.311	11.057 %	0.266
11202	482	0.270	0.000	0.000 %	0.274
11289	282	0.158	324.069	17.376 %	0.274
10228	175	0.098	163.990	11.429 %	0.278
11009	725	0.406	336.090	20.552 %	0.279
10881	1026	0.574	399.833	25.536 %	0.281
10609	892	0.499	285.194	14.574 %	0.282
11003	861	0.482	331.269	23.229 %	0.285
10932	1000	0.560	328.229	21.400 %	0.287
11019	568	0.318	204.465	21.127 %	0.289

then, require less time and less effort to be completed. They might also be brilliant individuals, being able to work alone without requiring leaders to control them. Peripheral workers are less stressed than other workers: in the assessment, even normal workers with a similar number of worked instances perform worse than peripheral workers. Stress may be fault of leadership [44] recalling all the time workers to their duties and judging commitment. Finally, they might be more free (in work) than other workers. Freedom in a workplace may conduct to a better working behaviour and satisfaction [45].

C. Finding weighted leaders

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 insights on how to improve it described in the "Conclusion and Future Work" section. The approach described in [9] (resumed 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 neighbor 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 resources. The workers with the greater number of process instances are considered to be leaders. The weight is 1 for the worker with 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 III. LEADER PROXIMITY AND MAXIMUM PROXIMITY TO OTHER WORKERS IN THE BPI CHALLENGE 2012 EVENT LOG.

Worker	Leader proximity	Max worker proximity
10861	1.000	0.308
11181	0.972	0.327
11169	0.960	0.278
10913	0.932	0.308
10821	0.931	1.000
11119	0.931	0.300
11180	0.884	0.359
11254	0.884	1.000
10909	0.871	0.305
11203	0.855	0.315
11189	0.797	0.298
11201	0.792	0.367
10982	0.753	0.272
10125	0.697	1.000
11049	0.697	0.313
11259	0.613	0.314
11122	0.594	0.271
10899	0.578	0.302
11079	0.563	0.563
10124	0.513	0.600
11319	0.415	0.485
10914	0.393	0.444
11002	0.369	0.396
11299	0.366	0.428
11339	0.356	0.615
10939	0.345	0.345
10935	0.345	0.370
11000	0.338	0.348
11111	0.338	0.348
11269	0.333	0.333
10863	0.328	0.328
11121	0.324	0.324
10809	0.324	0.410
11302	0.323	0.378
11309	0.314	0.367
10972	0.313	0.449
10931	0.313	0.322
10933	0.310	0.363
11179	0.308	0.311
10929	0.302	0.311
11029	0.297	0.334
11019	0.289	0.310
10932	0.287	0.292
11003	0.285	0.301
10609	0.282	0.377
10881	0.281	0.289
11009	0.279	0.279
10228	0.278	0.286
11289	0.274	0.433
11202	0.274	0.282
10138	0.266	0.382
10910	0.265	0.265
10889	0.261	0.280
10789	0.261	0.377
10629	0.255	0.366
11001	0.248	0.267
10912	0.245	0.276
11300	0.244	0.274
11200	*0.234*	*0.240*
10880	*0.228*	*0.235*
10862	*0.221*	*0.250*
10859	*0.219*	*0.228*
10779	*0.215*	*0.231*
11304	*0.194*	*1.000*
10971	*0.192*	*0.192*
11120	*0.156*	*0.177*
10188	*0.098*	*0.126*

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 hyphotesis 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 the organization, or they have an expiring work contract so they try to work at their best to get a renewed contract.

Workers **10188** and **10779** are not generally involved much in the organization, registering a low number of worked instances and low collaboration with others. They may perform better in order to get more involved in the process (see [41]).

For workers **11200**, **10971** and **10862**, which do not fall in the previous categories, a possible explanation can be provided: the *job embeddedness* theory; they perform good in order to improve their organizational ties and to strenghten their position in the organization.

V. Conclusion and Future Work

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 responsability 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 responsabilities, 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

[1] R. Basu and S. G. Green, "Leader-member exchange and transformational leadership: An empirical examination of innovative behaviors in leader-member dyads," Journal of Applied Social Psychology, vol. 27, no. 6, pp. 477–499, 1997.

[2] D. J. Brass, "Being in the right place: A structural analysis of individual influence in an organization," Administrative Science Quarterly, pp. 518–539, 1984.

[3] M. E. Burkhardt and D. J. Brass, "Changing patterns or patterns of change: The effects of a change in technology on social network structure and power," Administrative science quarterly, pp. 104–127, 1990.

[4] M. E. Newman, "A measure of betweenness centrality based on random walks," Social networks, vol. 27, no. 1, pp. 39–54, 2005.

[5] L. C. Freeman, "A set of measures of centrality based on betweenness," Sociometry, pp. 35–41, 1977.

[6] M. Rostampoor-Vajari, "What is sociometry and how we can apply it in our life?" Advances in Asian Social Science, vol. 2, no. 4, pp. 570–573, 2012.

[7] J. A. Gruman and A. M. Saks, "Performance management and employee engagement," Human Resource Management Review, vol. 21, no. 2, pp. 123–136, 2011.

[8] R. Griffin and G. Moorhead, Organizational behavior. Cengage Learning, 2011.

[9] V. Blondel, C. De Kerchove, E. Huens, and P. Van Dooren, "Social leaders in graphs," Positive Systems, pp. 231–237, 2006.

[10] W. M. Van Der Aalst, H. A. Reijers, and M. Song, "Discovering social networks from event logs," Computer Supported Cooperative Work (CSCW), vol. 14, no. 6, pp. 549–593, 2005.

[11] V. der Aalst et al., "Business process mining: An industrial application," Information Systems, vol. 32, no. 5, pp. 713–732, 2007.

[12] J. M. Levine, H.-S. Choi, and R. L. Moreland, "Newcomer innovation in work teams," Group creativity: Innovation through collaboration, pp. 202–224, 2003.

[13] M. R. Leary and R. F. Baumeister, "The nature and function of self-esteem: Sociometer theory," Advances in experimental social psychology, vol. 32, pp. 1–62, 2000.

[14] W. Van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," Knowledge and Data Engineering, IEEE Transactions on, vol. 16, no. 9, pp. 1128–1142, 2004.

[15] B. F. van Dongen, A. K. A. de Medeiros, H. Verbeek, A. Weijters, and W. M. Van Der Aalst, "The prom framework: A new era in process mining tool support," in Applications and Theory of Petri Nets 2005. Springer, pp. 444–454, 2005.

[16] G. Daniela, "Business process intelligence," Computers in Industry, vol. 53, no. 3, pp. 321–343, 2004.

[17] P. J. Carrington, J. Scott, and S. Wasserman, Models and methods in social network analysis. Cambridge university press, vol. 28, 2005.

[18] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," PHYSICAL REVIEW E Phys Rev E, vol. 69, pp. 26–113, 2004.

[19] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," Nature, vol. 435, no. 7043, pp. 814–818, 2005.

[20] A. Lancichinetti, S. Fortunato, and J. Kertsz, "Detecting the overlapping and hierarchical community structure in complex networks," New Journal of Physics, vol. 11, no. 3, pp. 100–115, 2009.

[21] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment, vol. 2008, no. 10, pp. 201–213, 2008.

[22] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," PHYSICAL REVIEW E Phys Rev E, vol. 76, pp. 104–120, 2007.

[23] M. Ovelgönne and A. Geyer-Schulz, "An ensemble learning strategy for graph clustering," Graph Partitioning and Graph Clustering, vol. 588, p. 187, 2012.

[24] S. Liu, Q. Kang, J. An, and M. Zhou, "A weight-incorporated similarity-based clustering ensemble method," in Networking, Sensing and Control (ICNSC), 2014 IEEE 11th International Conference on. IEEE, pp. 719–724, 2014.

[25] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in Advances in Neural Information Processing Systems. MIT Press, pp. 849–856, 2001.

[26] F. D. Malliaros and M. Vazirgiannis, "Clustering and community detection in directed networks: A survey," Physics Reports, vol. 533, no. 4, pp. 95–142, 2013.

[27] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," Social Networks, vol. 32, no. 3, pp. 245–251, 2010.

[28] Y. Ding, E. Yan, A. Frazho, and J. Caverlee, "Pagerank for ranking authors in co-citation networks," Journal of the American Society for Information Science and Technology, vol. 60, no. 11, pp. 2229–2243, 2009.

[29] J. Baumes, M. Goldberg, and M. Magdon-Ismail, "Efficient identification of overlapping communities," in Intelligence and Security Informatics. Springer, pp. 27–36, 2005.

[30] W. J. Stevenson and M. Hojati, Operations management. McGraw-Hill/Irwin Boston, vol. 8, 2007.

[31] J. M. Levine, R. L. Moreland, and H.-S. Choi, "Group socialization and newcomer innovation," Blackwell handbook of social psychology: Group processes, vol. 3, pp. 86–106, 2001.

[32] T. Hansen and J. M. Levine, "Newcomers as change agents: Effects of newcomers' behavioral style and teams' performance optimism," Social Influence, vol. 4, no. 1, pp. 46–61, 2009.

[33] G. R. Jones, "Socialization tactics, self-efficacy, and newcomers' adjustments to organizations," Academy of Management journal, vol. 29, no. 2, pp. 262–279, 1986.

[34] D. C. Feldman, "Who's socializing whom? the impact of socializing newcomers on insiders, work groups, and organizations," Human Resource Management Review, vol. 4, no. 3, pp. 213–233, 1994.

[35] D. H. Gruenfeld and E. T. Fan, "What newcomers see and what oldtimers say: Discontinuities in knowledge exchange," Shared cognition in organizations: The management of knowledge, pp. 245–266, 1999.

[36] R. I. Sutton and M. R. Louis, "How selecting and socializing newcomers influences insiders," Human Resource Management, vol. 26, no. 3, pp. 347–361, 1987.

[37] H.-S. Choi and J. M. Levine, "Minority influence in work teams: The impact of newcomers," Journal of Experimental Social Psychology, vol. 40, no. 2, pp. 273–280, 2004.

[38] H.-S. Choi and L. Thompson, "Old wine in a new bottle: Impact of membership change on group creativity," Organizational Behavior and human decision processes, vol. 98, no. 2, pp. 121–132, 2005.

[39] N. L. Kerr and J. M. Levine, "The detection of social exclusion: Evolution and beyond." Group Dynamics: Theory, Research, and Practice, vol. 12, no. 1, p. 39, 2008.

[40] R. F. Baumeister, J. M. Twenge, and C. K. Nuss, "Effects of social exclusion on cognitive processes: anticipated aloneness reduces intelligent thought." Journal of personality and social psychology, vol. 83, no. 4, p. 817, 2002.

[41] J. K. Maner, C. N. DeWall, R. F. Baumeister, and M. Schaller, "Does social exclusion motivate interpersonal reconnection? resolving the" porcupine problem."," Journal of personality and social psychology, vol. 92, no. 1, p. 42, 2007.

[42] T. R. Mitchell, B. C. Holtom, T. W. Lee, C. J. Sablynski, and M. Erez, "Why people stay: Using job embeddedness to predict voluntary turnover," Academy of management journal, vol. 44, no. 6, pp. 1102–1121, 2001.

[43] I. R. Pinto, J. M. Marques, J. M. Levine, and D. Abrams, "Membership status and subjective group dynamics: Who triggers the black sheep effect?" Journal of personality and social psychology, vol. 99, no. 1, p. 107, 2010.

[44] L. R. Offermann and P. S. Hellmann, "Leadership behavior and subordinate stress: A 360" view." Journal of Occupational Health Psychology, vol. 1, no. 4, p. 382, 1996.

[45] A. O. Agho, C. W. Mueller, and J. L. Price, "Determinants of employee job satisfaction: An empirical test of a causal model," Human Relations, vol. 46, no. 8, pp. 1007–1027, 1993.