Check for updates

A Robust Process to Identify Pivots Inside Sub-communities in Social Networks

Joseph Ndong[1] and Ibrahima Gueye[2](✉)

[1] Department of Mathematics and Computer Science, University Cheikh Anta Diop, BP 5005 Fann Dakar, Senegal
joseph.ndong@ucad.edu.sn

[2] Computer Engineering and Telecom, Polytechnic School, Thies, Senegal
igueye@ept.sn

Abstract. In this work, we extend a previous work where we proposed a suitable state model built from a Karhunen-Loeve Transformation to build a new decision process from which, we can extract useful knowledge and information about the identified underlying sub-communities from an initial network. The aim of the method is to build a framework for a multi-level knowledge retrieval. Besides the capacity of the methodology to reduce the high dimensionality of the data, the new detection scheme is able to extract, from the sub-communities, the dense sub-groups with the definition and formulation of new quantities related to the notions of energy and co-energy. The energy of a node is defined as the rate of its participation to the set of activities while the notion of co-energy defines the rate of interaction/link between two nodes. These two important features are used to make each link weighted and bounded, so that we are able to perform a thorough refinement of the sub-community discovery. This study allows to perform a multi-level analysis by extracting information either per-link or per-intra-sub-community. As an improvement of this work, we define the notion of pivot to relate the node(s) with the greatest influence in the network. We propose the use of a thorough tool based on the formulation of the transformation of a suitable probabilistic model into a possibilistic model to extract these pivot(s) which are the nodes that control the evolution of the community.

Keywords: Social network analysis · Community detection · Energy Pivot · Influencer

1 Introduction

Social networks describe web-based services that allow users (individuals) to connect with other users, communicate, share or publish contents within the network [1,2]. The rise of web 2.0 has come with the ease in the production and sharing of content that allowed the social networks development. Users themselves become the producers of web content [1,3]. In a formal and simple description, social network can be seen as a graph consisting of nodes (as individuals) and links (as social links) used to represent social relations on social network sites [4]. These last years, Social networks sites have

C. M. F. Kebe et al. (Eds.): InterSol 2018, LNICST 249, pp. 248–260, 2018.
https://doi.org/10.1007/978-3-319-98878-8_23

become some important sources of online interactions, contents sharing [5] and communication means. While giving these opportunities, they are affordable and universally acclaimed. Social network sites are commonly known for information dissemination, personal activities posting [6], product reviews, online pictures sharing, professional profiling, advertisements, subjectivity [6], assessments [7], approaches [8], influences [9], observations [10], feelings [3], opinions and sentiments expressions [11], news, remarks, reactions, or some many other content type [12]. News alerts, breaking news, political debates and government policy are also posted and analyzed on social network sites.

Observations [13] show that more and more people are becoming interested and involved in social networks. Sometimes, users of social networks exploit them for making decisions. These decisions can be maked based on know or familiar users or, in some case, based on unfamiliar users [11]. Such a situation increases the degree of confidence in the credibility of these sites. The social network has transformed the way different entities procure and retrieve valuable information regardless of their location. The social network has also given users the privilege of giving their opinion.

The massive data generated or produced by all the social networks users allows for thorough analysis to efficiently extract useful knowledge such as trends, opinion leaders, influencers, feelings, etc. Such an analysis can be done as a whole, which means the representations of all the social network actors then proceed to the detection of communities and/or opinion leaders. The main aim of community detection methods is to partition the network into dense regions of the graph. Those dense regions typically correspond to entities which are closely related. The determination of such communities is useful in the context of a variety of applications such as customer segmentation, recommendations, link inference and influence analysis [14].

In this work, we focus on community influencers identification. We extend our previous work [15] in which, we derive a suitable state model from Karhunen-Loeve Transformation [16]. From this state model we built a new decision process, which allows us to extract useful knowledge and information about the identified underlying sub-communities from an initial network.

This extension is interested in the extraction of the most “important” nodes in given sub-communities already detected or known. Identifying sub-groups under a social network is a great challenge in the area of social network analysis [17–19], but analyzing the intrinsic behavior of each sub-group might be an important need if one desires to focus on the quality of nodes/actors [20]. If sub-communities are identified, decisions could be taken on a group of nodes regardless the underlying contents of the group itself since it can be viewed as a single homogeneous entity in which all items are of same behavior. One could also see the sub-community as a machine where the dynamic structure and the evolution depend strongly on the nature and quality of its combined pieces. If this second point-of-view is adopted, we need a tool to identify the “most important” pieces whose actions are more influential than those of the rest of the group. We define these pieces as “pivots” and the quality of the network is as important as it contains influential pivots with high degree of influence. In the following, we define pivots and how they could be important for a manager, a advertiser and so on.

Inside a sub-community, nodes share links so, identifying pivots can be helpful in many situations:

- (i) if in a sub-community, the pivot(s) is/are linked to a few other nodes, the manager can conclude that this network is of less quality simply because the "most" important nodes have less interactivity and thus less influence. But in the contrary, he can pay more attention to the underlying evolution of the group;
- (ii) the different sub-communities can be classified in an ascending order based on their quality – the first group is the one with the greater number of pivots, and so on;
- (iii) in a sub-community, we can find other sub-communities in the following manner: If the pivots themselves are not linked to each other, each of them can be the center of a group/entity formed by all the nodes linked to a pivot.

The procedure for sub-communities detection/identification we built in our previous work [15] consisted in extracting a decision variable from a state-space model we defined with a Karuhen-Loeve Transformation (KLT). This have served also to reducing the dimensionality of the dataset in order to maintain the only relevant part of the data. Then, after defining some new quantities as "energy" of an actor/node and the"co-energy between" actors, we apply a decision process to identify the sub-communities and the features inside each of them. The improvement we bring in this present work is the ability to learn more about the impact of node's energy to see if some nodes could be selected as those with the highest degree of influence. The notion of highest degree must be properly define to achieve this aim. We found that a suitable definition of a probability density function (pdf) related to the defined node's energy can be done. Then, a second definition of a possibilistic model will be able to discover the relevant node(s) we classify as pivot(s).

The rest of this paper is organized as follow: In Sect. 2 we do a thorough study of the related work; in Sect. 3 we detail the methodology and algorithm for the sub-community detection; in Sect. 4 we present our use of possibility theory to identify the pivots; in Sect. 5 we present the results of our experimental validation of our propositions, and we conclude this paper in Sect. 6.

2 Background

Social sites have undoubtedly bestowed unimaginable privilege on their users to access readily available never-ending uncensored information. Twitter, for example, permits its users to post events in real time way ahead the broadcast of such events on traditional news media. Also, social network allow users to express their views, be it positive or negative [20]. Organizations are now conscious of the significance of consumers' opinions posted on social network sites to the patronage of their products or services and the overall success of their organisations. On the other hand, important personalities such as celebrities and government officials are being conscious of how they are perceived on social network. These entities follow the activities on social network to keep abreast with how their audience reacts to issues that concerns them [21].

Opinion of influencers on social network is based largely on their personal views and cannot be hold as absolute fact. However, their opinions are capable of affecting the decisions of other users on diverse subject matters. For example, Rihanna's rant has cost Snapchat $800 million in one Day [22]. Opinions of influential users on Social network often count, resulting in opinion formation evolvement. Clustering technique of data mining can be used to model opinion formation by assessing the affected nodes and unaffected nodes. Users that depict the same opinion are linked under the same nodes and those with opposing opinion are linked in other nodes. This concept is referred to as homophily in social network [23]. Homophily can also be demonstrated using other criteria such as race and gender [24].

Researchers in social network analysis are facing many other research issues and challenges such as in *Linkage-based and Structural Analysis*, or in *Dynamic Analysis and Static Analysis*.

Linkage-based and Structural Analysis consists of analyzing of the linkage behaviour of the social network so as to ascertain relevant nodes, links, communities and imminent areas of the network - Aggarwal [20].

Static analysis, such as in bibliographic networks, is presumed to be easier to carry out than those in streaming networks. In static analysis, it is presumed that social network changes gradually over time and analysis on the entire network can be done in batch mode. Conversely, dynamic analysis of streaming networks like Facebook and YouTube are very difficult to carry out. Data on these networks are generated at high speed and capacity. Dynamic analysis of these networks are often in the area of interactions between entities such as in Papadopoulos et al. [25] and Sarr et al. [26]. Dynamic analysis is also covered through temporal events on social networks such as in Adedoyin-Olowe et al. [27] and Becker et al. [17], and we also have such analysis through the study of evolving communities - Fortunato [18], Sarr et al. [19].

3 Methodology and Algorithm for the Sub-communities Detection

This methodology tracks and detects sub-communities based on the analysis of a huge number of features corresponding to events/activities for which a group of actors/nodes participate. First, we aim at finding the main features, to incorporate in our model, by means of extended principal component analysis. The second relevant issue of this methodology is related to the specification of a new detection procedure consisting of merging all the relevant features into a single process we will label as a "Decision Variable" (DV). By analyzing this process for the sub-community tracking operation, we can discover subgroups of actors using a multi-level thresholding and the notion of "energy dissipation" of an actor over the events.

We consider a community of R actors $\Omega = (a_1, \ldots, a_R)$ which perform activities on a set of K initial correlated events $(e_1, \ldots, e_K)$. For each event e_k, we have a column vector of size R containing the amount of participation of all R actors to the corresponding activity. This operation gives us the $R \times K$ matrix of correlated random variables $X = (X_1, \ldots, X_K)$. In other words, one observes these random variables through R independent realization vectors $x^i = (x^i_1, \ldots, x^i_K)$ $i = 1, \ldots, R$.

After extracting the relevant components from the Karhunen-Loeve transformation [15], we can build our decision variable as a row vector $DV = (y_1, \ldots, y_K)$. Then we can set a certain number of concepts for our methodology. We introduce the notion of "energy dissipation" (Ed) to quantify the degree of importance a given actor puts on a series of events. This notion is simple and intuitive. When considering the set of events/activities, the events for which the actor puts a high degree of importance constitutes his energy. For example, we can consider money as energy. When someone goes to buy some products, we can say that he/she is dissipating a certain amount of his/her energy. In this case, he/she should buy a "product A", and consequently buy another "product B" necessary to use the product A. Here, we can see the notion of correlation between these products/variables. When an athlete performs several disciplinary exercises in sport, we can view his actions as the dissipation of his energy over the different events, in order to win a medal. The energy of an actor is then quantifiable, its a measure of the strength of his participation to the a series of activities.

If the actor participates actively to all or most of the activities with a high intensity, then his energy increases, otherwise we say that this actor has less energy according to the ensemble of events happening at a given period of time.

Since the DV variable contains the aggregated amount of all actors participation to all events, the energy dissipation Ed of an actor i is the row vector defined as:

$$Ed_i = \left\{k, / x_k^i \geq \text{DV}[k], \forall k = 1, \ldots, K\right\} \tag{1}$$

Ed_i contains all the index of events for which the energy dissipation is greater than the reference DV. Consequently, we can calculate the total energy of the actor i as:

$$E_i = \frac{|Ed_i|}{|DV|} \tag{2}$$

where $|.|$ indicates the size of a vector.

We also refer to the notion of "co-energy" dissipation (CED) as the amount of energy between two actors according to their participation to the same set of activities. This quantity is a measure of the mean energy produced simultaneously by the two actors on the same activities:

$$\text{CED}_{ij} = \frac{|\,(Ed_i \cap Ed_j)\,|}{|DV|} \tag{3}$$

Finally, our detection procedure boils down to fix a threshold α and put a link between actor i and actor j if the rate of their co-energy exceeds the limit α. This means the following inequality must be held to add the link:

$$\text{CED}_{ij} \geq \alpha \tag{4}$$

When Eq. 4 holds, the value of CED_{ij} becomes the weight of the link between actor i and actor j. And then, this link is bounded by the interval $[min(E_i, E_j), max(E_i, E_j)]$. By varying the threshold $\alpha \in [0; 1]$, one can build many different sub-communities with the same dataset, each sub-community with a score α which measures its degree of realization. The algorithm to achieve our aim is described as follow:

4 Using Possibility Theory to Identify Pivot(s)

Our pivot identification methodology relies mainly in the feature of energy we defined for the actors/nodes. Pivots refer to nodes with "high" energy. These nodes have the opportunity to control the dynamic evolution of the network. If the energy of a pivot decreases or increases, the structure of the network might evolve towards a new direction allowing to suppress or add links between nodes. A simple question arises from the perspective of pivots identification: how much energy is necessary for a node to be classified as a pivot? We believe that it is very difficult to answer this question by analyzing only the amount of energy of each node. If someone would like to do so, he/she should build a kind of threshold and apply the decision to put the label "pivot" on a node if its energy exceeds this limit. This methodology weakens the objectivity of pivots' identification. To surround this difficulty, we propose a more robust identification scheme based on a link between probability theory and possibility theory.

Algorithm 1. *Sub-Community Discovering*

Input : C_t, a community
$\Omega(C_t)$, the sets of actors within C_t
$x^i = (x_1^i, \ldots, x_K^i)$ the vector of participation of actor i
$DV = (dv_1, \ldots, dv_K)$ the decision variable
α, the link detection threshold
Output : **V**, a sub-community

`/* Calculate Co-Energy dissipation between actors and apply threshold to add link*/`
begin
  foreach $(k, l) \in \Omega(C_t)$, $k \neq l$ **do**
    `/* Apply Eq. (1)*/`
    $Ed_k = \{p, / x_p^k \geq dv_p, \forall p = 1, \ldots, K\}$
    $Ed_l = \{p, / x_p^l \geq dv_p, \forall p = 1, \ldots, K\}$
    `/* Apply Eq. (3)*/`
    $\text{CED}_{kl} = \frac{|(Ed_k \cap Ed_l)|}{|DV|}$
  `/*Apply threshold to decide to put a link, Eq. (4)*/`
  if $CED_{kl} \geq \alpha$ **then**
    $addLink(V, k, l)$
  return *V*

The energy property can be viewed as a continuous random variable $X : \Omega \rightarrow V$, where Ω is the set of actors/nodes and V the measurable function giving the real value of the energy as defined in Eq. 2. X does not return a probability. But we want to know, if it were the case, could this probability help achieving our goal.

Probability theory is a valuable quantitative tool to study randomness/uncertainty in random phenomena. In this area, we can find the probability of occurrence of different possible outcomes in an experiment. If a random variable returns a probability P, then

$V = [0, 1]$ and $\sum_{u} P(X = u) = 1$. As we defined the energy in Eq. 2, the energy E_i of actor i is always between 0 and 1. So, we can take $V = [0, 1]$, but we do not have $\sum_{i}^{K} E_i = 1$ (K is the number of nodes).

4.1 Defining Probability to Characterize Energy

For the purpose to return a probability from X, we have just to normalize the energy by the size of $\mathbb{E}$, where $\mathbb{E} = \{E_i, i = 1 \ldots, |\Omega|\}$. At this point, the random variable X returns a probability since the equation $\sum_{u} P(X = u) = 1$ holds. We build this probability to characterize each energy of a node with a probability of occurrence in $[0, 1]$. Now, using this probability definition, our scope is to build a test to identify the potential pivots under each sub-community. Generally, a typical test consists of applying a threshold on the outcomes (probabilities of energy) to decide to put the label pivot on a node if the probability distribution takes a value higher than this threshold. The task to build this threshold is not straightforward even if, for given nodes if the probability of their respective energy is 0.95, 0.6 and 0.3 for example, how could we decide to label a node as pivot, based only on the not obvious notion of "high" probability. Is 0.6 a "high" or "low" probability? By pointing out this example, we simply want to show that it is not evident to know the "best" value of the probability threshold in order to conclude if the node is a pivot or not. Nevertheless, one could, for simplicity, build an heuristic decision process where the threshold is set manually. With probability, the only evident decision we can take is, when the probability of a node is $P(X = u) = 1$. If this case happens, it means that there's only one pivot in the entire sub-community, since the other nodes have probability 0 and so this sub-community is built with only one node. Finally, we believe that pivots might be nodes with any probability, different to zero, "sufficient" to become member of a sub-community and to have the potential to change the dynamic evolution of the network. To go towards the direction of extracting the "best" level of probability measure, we think of an alternative related the area of possibility theory, which gives as another tool to represent uncertainty in a qualitative fashion. This tool can also helps to learn more about the **incompleteness** that reflects the lack of information. We have seen above that, our defined probability doesn't give us the information about the limit to apply to detect pivots. So, the idea behind the use of this new scheme is to associate to each node, in accordance to its energy, a degree of possibility which quantifies the level of "importance" of that node among the others.

4.2 Detecting Pivot by Possibility Degree

Between probability and possibility, we can state a consistent principle is this terms: "*what is probable should be possible*" [28]. This requirement can then be translated as follow:

$$P(A) \leq \Pi(A) \qquad \forall A \subseteq \Omega \tag{5}$$

where P and Π are, respectively, the probability and the possibility measure on the domain Ω. In this case, Π is said to **dominate** P. With Π, given nodes should have the maximum possibility degree, i.e. the value 1. In possibility theory the equality $\sum_u \Pi(X = u) = 1$ is not guaranteed. So, for any node, when its possibility degree reaches the maximum, we can robustly say that this node is a pivot since it is entirely sure that it might exist in the network. So, we see that the notion of "high" probability can be defined properly, since it corresponds to the maximum degree of possibility.

A possibility measure, [29], Π on V is characterized by a possibility distribution $\pi : V \rightarrow [0, 1]$, and is defined by:

$$\forall A \subseteq V, \Pi(A) = sup\{\pi(v), v \in A\}. \tag{6}$$

If V is a finite set, thus: $\forall A \subseteq V, \Pi(A) = max\{\pi(v), v \in A\}$. The key concept of a possibility distribution is the preference ordering it establishes on V. Basically, π designates what one knows about the value of a given variable X, and $\pi(v) > \pi(v')$ states that $X = v$ is more plausible than $X = v'$. When $\pi(v) = 0$, thus, v is an impossible value of the variable X while $\pi(v) = 1$ means that v is one of the most plausible values of X. For us, identifying pivots is just a process to searching at these most plausible values.

Transforming a probability measure into a possibilistic one then amounts to choosing a possibility measure in the set $\Im(P)$ of possibility measures dominating P. This should be done by adding a strong order preservation constraint, which ensures the preservation of the shape of the distribution:

$$p_i < p_j \Leftrightarrow \pi_i < \pi_j \qquad \forall i, j \in \{1, \ldots, q\}, \tag{7}$$

where $p_i = P(\{E_i\})$ and $\pi_i = \Pi(\{E_i\}), \forall i \in \{1, \ldots, K\}$. It is possible to search for the most specific possibility distribution verifying (5) and (7). The solution of this problem exists, is unique and can be described as follows. One can define a strict partial order P on Ω represented by a set of compatible linear extensions $\Lambda(\mathsf{P}) = \{l_u, u = 1, L\}$. To each possible linear order l_u, one can associate a permutation σ_u of the set $\{1, \ldots, q\}$ such that:

$$\sigma_u(i) < \sigma_u(j) \Leftrightarrow (\omega_{\sigma_u(i)}, \omega_{\sigma_u(j)}) \in l_u, \tag{8}$$

The most specific possibility distribution, compatible with the probability distribution $(p_1, p_2, \ldots, p_K)$ can then be obtained by taking the maximum over all possible permutations:

$$\pi_i = \max_{u=1,L} \sum_{\{j | \sigma_u^{-1}(j) \leq \sigma_u^{-1}(i)\}} p_j \tag{9}$$

Finally, the vector $(\pi_1, \pi_2, \ldots, \pi_K)$ gives us all the possibility degrees for the K nodes, corresponding to the $(p_1, p_2, \ldots, p_K)$ vector of probability of their energies. And pivots are nodes for which the possibility degree exceeds a given rate δ. For this study we set this threshold $\delta = 1$, it corresponds to the maximum value a possibility degree might be set. A more flexible and non-heuristic method might be to set δ to a value less than the maximum. But, here, we set δ to 1 in order to show that in all situations or scenario, a pivot must be found.

For a thorough view of possibility theory, we recommend the reader to [28–30].

5 Validation

We validate our approach on the real world collection of data coming from **Reddit.com** [31]. We use several samples of different sizes and, build four scenarios A, B, C and D with dimension ($N \times K$, N the number of actors and K the number of events) 10×15, 10×150, 10×500 and 10×1200 respectively. In Table 1, we give an idea on the content of the data, in each column vector, we have the total amount of submissions to an image by the set of actors.

Table 1. Activities and amount of actor participation to submissions on events. Scenario A.

Actors	Events														
	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}	e_{12}	e_{13}	e_{14}	e_{15}
1	11	0	11	4	0	2	0	4	18	2	0	6	16	1	0
2	5	0	0	0	0	0	1	0	0	0	2	0	2	0	0
3	1	0	2	0	3	1	1	0	1	0	0	2	1	0	2
4	4	1	0	0	0	1	2	0	7	0	1	0	9	1	0
5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
7	0	2	0	0	1	0	0	0	0	0	0	4	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

In the previous experiment our previous work [15] (that we can not show again due to a limit page number), we retrieved two levels of information can be retrieved from the results. In the first level, we had the results about the formation of the underlying sub-communities. It corresponds to the natural clustering of the different nodes according to the energy provided by each of them. The second level of information refered to the characteristics of links and nodes inside the given sub-groups. This refinement provides useful information when one wants to emphasize and explore some parts of the network.

Here in below, we discuss the obtained results about the pivots identification.

Discussion About the Pivots Identification Results

We apply the detection process to identify pivots to the data samples for scenarios A, B, C, D, E, F and G. We have arbitrarily chosen the size of the sample and the interval where data come from. We just want to show cases where there's one pivot or more. The Table 2 resumes the data selection scheme, and in we put the results of the detection procedure in Table 3 where the variable NBE refers to the total amount of participation on events by nodes, p is the probability of the energy and π the corresponding degree of possibility. The results give many useful information. For the scenarios A, B, C and D, we extract only one pivot which have the particularity to be the node with the highest amount of activities in the network. The pivot also has the highest probability of energy

Table 2. Information about the selection of the different data samples

Scenario	A	B	C	D	E	F	G
Size	15	150	500	1200	155	255	555
$interval$	[1:15]	[1:150]	[1:500]	[1:1200]	[1100:1254]	[1000:1254]	[700:1254]

Table 3. Detection of pivots when degree of possibility $\pi = 1$

Nodes i	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	Scenario
NBE	75	10	14	26	1	2	7	0	0	0	A
p	0.30	0.10	0.25	0.25	0	0	0.10	0	0	0	
π	1	0.20	0.70	0.70	0	0	0.20	0	0	0	
pivot	Yes	No	No	No	No	No	No	No	No	No	
NBE	591	28	218	142	44	31	117	53	10	8	B
p	0.42	0.06	0.18	0.23	0	0	0.11	0	0	0	
π	1	0.06	0.35	0.59	0	0	0.17	0	0	0	
pivot	Yes	No	No	No	No	No	No	No	No	No	
NBE	1638	108	668	348	120	63	348	170	27	16	C
p	0.39	0.11	0.17	0.22	0	0	0.11	0	0	0	
π	1	0.22	0.38	0.61	0	0	0.22	0	0	0	
pivot	Yes	No	No	No	No	No	No	No	No	No	
NBE	3659	257	1596	773	206	164	890	425	59	40	D
p	0.39	0.11	0.17	0.22	0	0	0.11	0	0	0	
π	1	0.22	0.38	0.61	0	0	0.22	0	0	0	
pivot	Yes	No	No	No	No	No	No	No	No	No	
NBE	428	34	204	95	26	40	52	40	6	4	E
p	0.38	0	0.38	0	0.06	0.06	0.06	0.06	0	0	
π	1	0	1	0	0.25	0.25	0.25	0.25	0	0	
pivot	Yes	No	Yes	No	No	No	No	No	No	No	
NBE	702	50	321	140	38	54	122	82	13	8	F
p	0.32	0	0.32	0.07	0	0	0.07	0.15	0	0.07	
π	1	0	1	0.23	0	0	0.23	0.38	0	0.23	
pivot	Yes	No	Yes	No	No	No	No	No	No	No	
NBE	1569	127	663	335	72	81	405	170	22	16	G
p	0.32	0	0.38	0	0.06	0	0.18	0.06	0	0	
π	0.62	0	1	0	0.12	0	0.31	0.12	0	0	
pivot	No	No	Yes	No	No	No	No	No	No	No	

but, the value of this probability seems to be "very low" in a pure point-of-view of the probability theory. For example, for scenario A, the pivot has a probability $p = 0.30$ but its possibility degree reaches $\pi = 1$. For scenarios A, B, C and D, we find only one pivot and it corresponds to the node with the highest probability and which have perform the greater amount of activities in the network. But this findings are not a generality since we see, in the scenario G, the pivot has the highest probability but it does not have the greatest amount of activities. This discover says clearly that, learning only the amount of activities is not a sufficient process to analyze a network in order to detect links or to put a level of importance/quality to nodes. In scenarios E and F, we detect twos pivots which have not the same amount of actions but they have the same probability of energy. We think that this situation is do to the fact that the quality of node depends not only to its level of participation on events but, it depends also to its interaction with other nodes. So one must have two (or more) nodes with different level of actions on events but they have the same behavior regarding to other nodes.

6 Conclusion

In this work, we have extent a new technique related to an extended version of principal component analysis to build a methodology for the purpose of community detection in a social network. The initial work built a technique more elaborated to run within stochastic process than the classical PCA which is designed originally to solve the problem of dimensionality reduction for univariate dataset. The main innovation of the work is manifold: (i) we define the notion of co-energy between two nodes to quantify the intensity of their relation, (ii) we can also extract the proper energy of a given node to know how it influences the overall community, (iii) technically, the KL-PCA technique makes possible to build a decision variable and to form a state model from which we apply a decision process to identify each link. The introduction of the notion of energy make possible to see potential intra sub-communities (i.e. nodes with the same co-energy) inside a sub-community; (iv) each detected link is bounded, so we know how much energy is necessary to maintain a link over time. In this complementary study, we show that a possibility distribution can be properly defined from the energy to solve an interesting feature i.e, the problem of identifying pivots which have the main impact in the dynamic nature of the network. From this work, we plane to learn the impact of the number of pivots in a given sub-community and between sub-communities to face the idea related to their impact on a network distributed in many geographic area.

References

1. Adedoyin-Olowe, M., Gaber, M.M., Stahl, F.T.: A survey of data mining techniques for social media analysis. JDMDH **2014** (2014)
2. Chen, Z., Kalashnikov, D.V., Mehrotra, S.: Exploiting context analysis for combining multiple entity resolution systems. In: ACM SIGMOD International Conference on Management of Data, SIGMOD 2009, pp. 207–218 (2009)
3. Kaplan, A.M., Haenlein, M.: Users of the world, unite! the challenges and opportunities of social media. Bus. Horizons **53**(1), 59–68 (2010)
4. Borgatti, S.P.: Social network analysis, two-mode concepts in. In: Meyers, R.A. (ed.) Computational Complexity, pp. 2912–2924. Springer, Heidelberg (2012). https://doi.org/10.1007/978-1-4614-1800-9_179

5. Chelmis, C., Prasanna, V.K.: Social networking analysis: a state of the art and the effect of semantics. In: IEEE International Conference on Social Computing, pp. 531–536 (2011)
6. Asur, S., Huberman, B.A.: Predicting the future with social media. In: Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology, pp. 492–499. IEEE Computer Society (2010)
7. Kim, Y., Hsu, S.-H., de Zúñiga, H.G.: Influence of social media use on discussion network heterogeneity and civic engagement: the moderating role of personality traits. J. Commun. **63**(3), 498–516 (2013)
8. Korda, H., Itani, Z.: Harnessing social media for health promotion and behavior change. Health Promot. Practice **14**(1), 15–23 (2013)
9. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Identifying 'influencers' on Twitter. In: Fourth ACM International Conference on Web Search and Data Mining (WSDM) (2011)
10. Atkinson, N., Chou, W.-Y.S., Hunt, Y.M., Beckjord, E.B., Moser, R.P., Hesse, B.W.: Social media use in the United States: implications for health communication. J. Med. Internet Res. (2009)
11. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (2008)
12. Liu, B.: Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, San Rafael (2012)
13. Statista: Number of social media users worldwide from 2010 to 2021 (in billions) (2017). https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/. Accessed 09 Oct 2017
14. Khatoon, M., Banu, A.: A survey on community detection methods in social networks. I. J. Educ. Manag. Eng. **1**, 8–18 (2015). https://doi.org/10.5815/ijeme.2015.01.02
15. Ndong, J., Gueye, I.: A new decision technique for sub-community and multi-level knowledge extraction in social networks. In: Cherifi, H., Gaito, S., Quattrociocchi, W., Sala, A. (eds.) Complex Networks & Their Applications V. Complex Networks 2016. SCI, vol. 693, pp. 263–274. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-50901-3_21
16. Gray, R.M., Davisson, L.D.: An Introduction to Statistical Signal Processing. Cambridge University Press, Cambridge (2004)
17. Becker, H., Naaman, M., Gravano, L.: Beyond trending topics: real-world event identification on Twitter. In: Proceedings of the Fifth International Conference on Weblogs and Social Media (2011)
18. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3), 75–174 (2010)
19. Sarr, I., Ndong, J., Missaoui, R.: Overlaying social networks of different perspectives for inter-network community evolution. In: Missaoui, R., Sarr, I. (eds.) Social Network Analysis - Community Detection and Evolution. LNSN, pp. 45–70. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12188-8_3
20. Aggarwal, C.C.: An introduction to social network data analytics. In: Aggarwal, C. (ed.) Social Network Data Analytics, pp. 1–15 (2011). https://doi.org/10.1007/978-1-4419-8462-3_1
21. Castellanos, M., et al.: LCI: a social channel analysis platform for live customer intelligence. In: ACM SIGMOD International Conference on Management of Data, pp. 1049–1058. ACM (2011)
22. FR, Y.N.: Rihanna's rant has cost Snapchat $ 800 million (2018). https://fr.news.yahoo.com/coup-gueule-rihanna-fait-perdre-114958692.html. Accessed 19 Mar 2018
23. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. Ann. Rev. Sociol. **27**(1), 415–444 (2001)
24. Jackson, M.O.: Social and Economic Networks. Princeton University Press, Princeton (2008)
25. Papadopoulos, S., Kompatsiaris, Y., Vakali, A., Spyridonos, P.: Community detection in social media. Data Min. Knowl. Discov. **24**(3), 515–554 (2012)

26. Sarr, I., Missaoui, R.: Temporal analysis on static and dynamic social networks topologies. In: Alhajj, R., Rokne, J. (eds.) Encyclopedia of Social Network Analysis and Mining. Springer, Heidelberg (2014). https://doi.org/10.1007/978-1-4614-6170-8
27. Adedoyin-Olowe, M., Gaber, M.M., Stahl, F.: TRCM: a methodology for temporal analysis of evolving concepts in Twitter. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2013. LNCS (LNAI), vol. 7895, pp. 135–145. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38610-7_13
28. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. In: Fuzzy Sets and Systems, pp. 3–28 (1978)
29. Dubois, D., Prade, H., Sandri, S.: On possibility/probability transformations. In: Proceedings of the Fourth International Fuzzy Systems Association World Congress (IFSA 1991), Brussels, Belgium, pp. 50–53 (1991)
30. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
31. Gueye, I., Ndong, J., Sarr, I.: An accurate probabilistic model for community evolution analysis in social network. In: International Conference on Signal-Image Technology Internet-Based Systems (SITIS), pp. 343–349 (2015)
32. McGloin, J.M., Kirk, D.S.: An overview of social network analysis. J. Crim. Justice Educ. **21**(2), 169–181 (2010)
33. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications, vol. 8. Cambridge University Press, Cambridge (1994)