

An Accurate Probabilistic Model for Community Evolution Analysis in Social Network

Ibrahima Gueye
Department of Mathematics and CS
University Cheikh Anta Diop
Dakar, Sénégal
ibrahima82.gueye@ucad.edu.sn

Joseph Ndong
Department of Mathematics and CS
University Cheikh Anta Diop
Dakar, Sénégal
joseph.ndong@ucad.edu.sn

Idrissa Sarr
Department of Mathematics and CS
University Cheikh Anta Diop
Dakar, Sénégal
idrissa.sarr@ucad.edu.sn

Abstract—The scope of the work is to build a framework able to study the evolution of a set of communities based on their underlying social activities. Generally, for a given community, many subgroups may exist and evolve with different and various opinions or behaviors. So, in this paper, we will focus on the identification of the potential subgroups and their potential relation/correlation corresponding to self-similarity over time. Clearly, we want to know if the subgroups remain unchanged then being stable or might they evolve to merge by forming new groups. In this respect, social engagement that refers to the participation of actors from a community to the activities of a social group is used to distribute activities into several classes. So building subgroups will be our first challenge and analyzing temporal correlation between them will be another interesting issue in this present work. The first problem can be solved by analyzing the activities inside the given initial community. We believe that, in many situations, activities should be characterized by parametric distributions as the gaussians. So, by means of the gaussian mixture modeler (GMM), subgroups can be identified successfully. Thereafter, the intrinsic relation between subgroups and their temporal evolution can be studied clearly with the calibration of hidden Markov models (HMM). The achievement of this study can help management operators to take decisions in two ways: i) since each GMM subgroup may correspond to a single individu's opinion/behavior, typical decision could be made for a given social group ii) also, the manager can take advantageous decisions by merging opinions for subgroups which have self-similarities; the HMM is here to learn more about this issue. We show the effectiveness of our approach by using real life data from Reddit.com.

Keywords—Social network analysis, community detection, Group cohesion, GMM, HMM, Viterbi;

I. INTRODUCTION

The analysis of social networks by focusing on community detection has attracted the attention of many researchers because of their importance and application domain [16], [11], [8], [9], [3], [24]. In fact, studying communities allows us to undermine and analyze the components of networks and to understand their patterns. For example, identifying the group of users who are interested in cinema allows better gauge the target of advertising a new film, and thus, avoid contacting individuals who would not be interested. A well-known approach for community detection is described in [7] and is based on the intuition that groups within a network may be detected through natural divisions among the vertices without requiring to set

the number of groups or put restrictions on the sample size. Many other approaches have been developed to tracking the evolution of social communities over time [2], [17], [11], [19]. Indeed, they use several static views of the network at different time instants. For each view, one may use an existing community detection algorithm [6] to depict the community topology. Therefore, between two time points, changes may occur such as a network growth or partition. Most of the community detection/evolution approaches are devised based on the observation that a community is a group of nodes with many connections among the group's members, but few to the rest of the network [16], [11], [17]. That is, the main feature to disclose a community is the density within the group's members. However, dense communities are one kind of group structure in networks, and there may be obviously many other features or structures that could help to understand a network. Recently, some works have focused on community evolution/detection by relying entirely on the group's members behavior with respect to the activities that occur in the network instead of using exclusively links and the density of the groups [9], [21], [22].

In our best of knowledge, none of the existing approaches has tracked the evolution and relation of the different groups while focusing on their temporal correlation/dependencies. We believe that potential time-dependency might exist between members of given subgroups. So discovering subgroups is in fact an interesting issue since it gauges how members of a group are close to each other. But also learning about temporal correlation between subgroups might be another promising knowledge since it helps to predict whether a community will remain stable over time or may be split/merge according to events happening on other groups. Another important fact is the possible relationships between groups. Learning temporal dependencies may be attached to the notion of "cohesion" of a group. Cohesion can be defined as the tendency for a group to be atomic while working towards a goal or to satisfy the emotional needs of its members [4]. Many dimensions are tied to group cohesiveness among which, we point out the dynamic nature, the instrumental basis, and the emotional dimension [4]. The dynamic nature refers to how cohesiveness gradually changes over time in terms of strength and shape. The instrumental basis refers to how people cohere whatever the

reasons. Finally, emotional dimension refers to how cohesion is pleasing to its group members.

There is a abundant literature pointing out group cohesiveness [24], [5], [23] but they usually rely on the network topology and/or are conducted to assess their correlation with the performances of a group.

Surprisingly, the dynamic nature of a group cohesiveness, which is tightly correlated to the group structure evolution, is less studied. This work proposes to track the dynamic of a community and therefore to know how its cohesiveness behaves over time. In other words, we aim at answering if disjoint subgroups can be found in a given community, which means that the cohesiveness is weak. To this end, we start from the hypothesis that members of a group should maintain their social engagement to cope with the emotional and instrumental basis of the group cohesion. That is, each member should take part to activities happening in the group. Prohaska, Anderson and Binstock highlighted that the term social engagement is commonly used to refer to an individual's participation in the activities of a social group [18]. The term was defined as "the extent to which an individual participates in a broad range of social roles and relationships" [1] and/or as "the commitment of a member to stay in the group and interact with other members" [25]. Moreover, social engagement is different from the concept of a social network, as social engagement focuses on an activity, rather than the group. Therefore, if we come up to show that individuals of a group do not have the same social engagement for all activities, we conclude that the group cohesiveness is not strong and we may therefore unveil the cohesive subgroups it contains.

To reach our goal, we assume that the social engagement of individuals with respect to a set of activities is a gaussian process. That is, there are some activities that attract more than others thanks to their importance or their nature. Hence, by gathering activities based on their nature or importance, we may observe a mixture of gaussian processes. Each normal distribution captures the view of a group of individuals as their social engagement to a set of activities. That is, the Gaussian mixture Model help detect the existence of subgroups within a community over time, which means that the overall community is not cohesive regarding to our previous definition. Thereafter, we learn more about these groups in order to find temporal correlations. So groups which have some amount of similarities will be grouped together to form what we means as super-group. This scope can be successfully achieved with the help of hidden Markov model. We have already used both GMM and HMM for the scope of anomaly detection, see [15].

The rest of this paper is structured as follows. In Section II, we talk about the reasons which allow us to learn temporal correlations. The architecture of our community tracking system is described in Section III where we explain how to discover initial subgroups by means of GMM framework and how we build super-groups over the subgroups by means of of HMM and Viterbi decoder. In Section IV we validate our approach with a real-world collection of dataset and conclude in Section V.

II. EXPLORING TEMPORAL DEPENDENCIES

Given a family of subgroups, we focus on learning more about the potential time correlations within these subgroups. In fact, we believe that elements of different subgroups might have "affinities" i.e they might have possible relationships. Another idea is the possibility for elements of different subgroups to have the same view or behavior with the arrival of new activities. So temporal correlations/dependencies is the fact that elements of a given subgroup might share some degree of relationship to other elements in other subgroups. The literature is full of techniques able to form classes from an initial given dataset from a network or community. For the scope of anomaly detection, we have already develop robust unsupervised clustering approach coming from Bayesian inference (as hierarchical clustering, method using exponentially distributions GMM) or based on fuzzy-possibilistic clustering. Techniques which rely on the fuzzy logic theory develop the property of membership degree; it means that every data point has a "probability" to belong in all classes. One has just to put this point in the class with the high degree. Recently the possibility theory develops some aspects that follow this principle. We have already used these techniques for the scope of anomaly detection (see [13], [14]) and also in social network analysis [22]. Despite all these approaches, GMM clustering reveals a strong capacity to classify data into a set of a priori unknown number of hidden classes. In nature, many processes develop the natural behavior to follow a normal distribution. We believe that the entire process which governs the network under consideration is an ensemble of gaussian components. So using GMM to perform the unsupervised clustering becomes a natural choice. When using the GMM framework, we resolve two major problems we interest in. First, we can discover easily the different subgroups from our initial community. Second, the GMM will give us the variance of each subgroup, telling us how the data inside a given class is spread around the mean. This variance could be very low for a cluster, so making the underlying elements of that group very close to each other. Consequently the given group becomes dense and potentially disjointed to the others. If the variance is high enough, it means that the corresponding elements are far from the mean of their class. We think that this situation might evolve to make correlations between these elements and members of the remaining groups with almost the same tendency or property. Since the different subgroups come from the same network and evolve in time, these potential correlations should be "temporal". If someone wants to learn correlation between subgroups coming from heterogeneous areas, then he/she is in presence of spatial correlations. This is not the scope of the study.

If we calibrate our GMM in a convenient way, we could study these temporal dependencies by means of hidden markov model, which describes another category of framework also based on exponential family of distributions. Simply, with the HMM, we can form different hidden states, each of them containing part of some given subgroups. Finally, each

state with its subgroups forms a new subgroup. Based on its transition matrix, the HMM gives us the possibility to calculate the probability for any given data point (element of a given subgroup) to reach the other subgroups. Based also on the emission probability, we can find how each data point belongs to a new group (state of the HMM). So if we plug into the HMM framework the previously found GMM subgroups, we can discover new families of members potentially coming from heterogeneous sub-groups. The HMM has the ability to discover many possible sequences of states where it matches the subgroups. By using the Viterbi decoder, we can find the single best sequence of states, which is optimal to capture the data for each subgroup. In the following, each state is labeled as a "super-group".

III. ARCHITECTURE OF THE COMMUNITY TRACKING SYSTEM

The main idea of this paper is related to the notion of **"activity importance"** that states the number of times actors of a given community participate to an activity. That is, for a given activity, its importance is the total number of participating actors. So actors which make an activity "important" might form a subgroup. Consequently, a given actor could impact several activities to make their "importance" increasing. So this actor will have a high scores to belong to many subgroups. Therefore, temporal dependencies might appear between these subgroups and we think they can be put in a same super-group.

One can study this feature for many reasons. The management of social communities might rely of the capacities of human expert to extract knowledge about actor's preferences about their participation to activities. For example one can need to discover groups of actors for a category of activities, or one can need to discover the more relevant activities for econometric, academic organization, etc. When considering a community, actors who perform a variety of activities might have different behaviors according to the scope and the importance of each activity. Thus, it is important to analyze the amount of reactions/participation on activities to classify the actors who have almost the same view on the system.

In this work, we assume that the activities under considerations follow processes which are normally distributed (gaussian). This assumption is intuitive since in nature, many processes are distributed with normal shapes. Since we assume also that the actors behaviors are heterogeneous, the importance activity can lead to a mixture of gaussian processes. Each normal distribution captures the view of a group of actors as their reactions to participate to a set of activities. We propose the system described in figure 1 to model the importance activity for the scope of building super-groups from an evolving community. There's several advantages to use this model. The figure 1 shows that the framework of GMM can be used to deal with the classification of the activities into some gaussian components. Thereafter we run a HMM to find temporal correlations between subgroups. This unsupervised clustering scheme gives generally two kinds

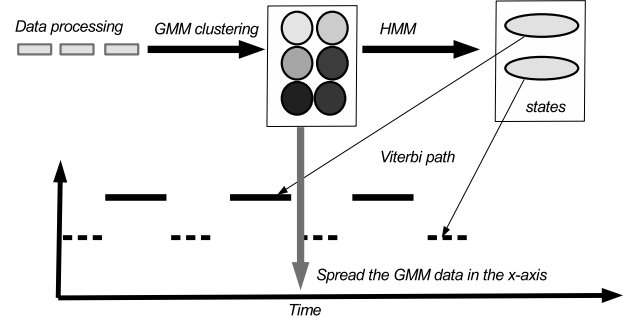


Fig. 1. Architecture of the system. The role of the GMM framework is to process the data from the initial network to build subgroups. Temporal correlations might happen between these subgroups. With the help of the HMM framework, we build super-groups, each covering part of the subgroups.

of partitions. In the first kind, a partition could be formed with well-formed and disjoint/independent clusters, whereas in the second one, a partition is composed of embraced and correlated clusters. The latter case could happen when the variance of the data is important or when the data have features of self-similarity. So, after running this operation, one can ask to know how the gaussian components are linked in time, i.e., do they exist correlation/dependencies of these distributions. The answer can be found by the tool of hidden markov model which make possible to build a set of a priori unknown hidden states where the gaussians are redistributed. By reusing the HMM trellis, we calibrate the Viterbi algorithm, to find the single best path showing how the HMM states evolve over time. This operation has the great advantage to retrieve the maximum likelihood to discover how the dataset (activities) is spread on the time scale and how the variation levels of the data are captured by the states. In other words, the time scale is divided in different time-intervals where data are spread over the states. Finally, in each time-interval, we can find the relevant activities and/or the actors which form a sub-community.

IV. MODEL EVALUATION

A. Experimental Data Collection

To validate our approach, we use a real world collection of data coming from **Reddit.com**. The dataset is a collection of 132,307 *reddit.com* submissions [10]. The main feature, for this data, is based on voting/submitting for "images". And for each image, re-submission is allowed. The (re)-submission operation is done by a given group (community) that we consider as a single actor based on the semantic we associate to the reaction each group has regarding to an image. In fact, even if the members of a group submit different votes, only the overall score representing the opinion of the all group is considered through this work. In this data set, the notion of "activity" is related to "submission" or "re-submission" to an image. Hence, the data set are quite the same one can get from Facebook regarding to posts/tags and users who

react to them. Moreover, since re-submission can be done, this dataset is particularly interesting based on the fact it clearly emphasizes the importance of the amount of activities an actor might have amount others. It is worth noting that the importance of activities is a predominant factor that might influence considerably communities creation.

By mining the huge number of activities within the dataset, one can find potential links between the different actors. The original dataset is composed of a set of 13 features ($\#image_id$, $\#unixtime$, $\#rawtime$, $\#title$, $\#total_votes$, $\#reddit_id$, $\#upvotes$, $\#subreddit$, $\#downvotes$, $\#localtime$, $\#score$, $\#comments$, $\#username$). The images are represented by the variable $\#image_id$ and actors which perform submissions and re-submissions are represented by the variable $\#subreddit$. Our aim is to analyze the amount of activities represented by the variable $\#total_votes$ to identify links between actors who perform them. This dataset is organized as a simple time-series vector. For each time step t we have a number corresponding to a amount of time an actor participate to an activity.

B. Discovering Subgroups with the GMM

In many research areas, unsupervised classification has become a valuable tool to propagate data into a priori unknown set of clusters. This operation has the ability to organize the dataset into different groups in order to learn more efficiently the dynamics of the process that generates the data, i.e., the activities. Our motivation to use gaussian mixture model (GMM) is based on our belief that the real distribution of this process (submitting/re-submitting to an activity) is an ensemble (mixture) of K gaussian components. Each of these normal distribution captures a part of the process, thus corresponding to a subset of the data sample.

Since, we deal with an unsupervised clustering operation, it is crucial to find the best value of K . In the GMM calibration procedure, each normal component is characterized by its mean and variance. So to know the best number of cluster, we calibrate a set of $r = \{2, 3, 4, \dots, K\}$ models. Finally, we choose the model that minimizes the variance and set the value K consequently.

C. Super-groups tracking with HMM

In the following, we deal with a two-step analysis using hidden Markov model to build super-groups over the K GMM subgroups previously discovered. A HMM can be used to solve three different problems. In the first case, it treats the evaluation problem of calculating $P(O|\lambda)$, i.e the probability of observing the data sequence O given a model λ . In the second problem, it takes a sequence O_T and try to find a state sequence I_T which is "optimal" in some sense; this is a typical estimation problem since it wants to uncover a part of the model. The third problem tries to adjust the model parameters in order to maximize $P(O|\lambda)$. The first and second problems are solved in this work to achieve our aim. To derive them, we use respectively the forward-backward (expectation-maximization-EM) algorithm and the Viterbi decoder. The

classical mathematical routines to implement the Viterbi algorithm by the expectation-maximization (EM) approach are in [12], [20].

Calibrating this HMM boils down to find P hidden states (super-groups) and redistribute the K GMM clusters. The literature is full of HMM types. In the work, we put the attention to discrete-time ergodic model which relies on the assumption that the probability to reach a state i given that we are in the same state is not equal to zero. We begin by transforming each cluster into a sequence of a finite set of alphabet (symbols 1,2,3,...), using a *maximum a posteriori* criteria. This discrimination phase will help for plugging the above subgroups into different a priori unknown states, using hidden markov model. So, due to the potential correlations between subgroups with the same probabilistic features, they are grouped into different super-groups. We represent these family of states by the following collection of unknown random variables $\{Q_1, Q_2, \dots, Q_T\}$ (where Q_t is a constant value with values in $\{1, 2, \dots, K\}$). We also represent our discrete alphabet by the vector $\{O_1, O_2, \dots, O_N\}$. Finally, the framework of HMM can be used to produce these states and to determine the probability of each symbol being in a state.

As we did for the calibration of the GMM, we need to find the best number of HMM states, since this framework acts a finite state machine which can do unsupervised clustering. We build a set of $p = \{2, 3, 4, \dots\}$ models and choose the one for which the different states are well separated. The elements of the main diagonal of the transition matrix for each model is analyzed to do the choice. All the models where the probability of staying in a state is less than 1/2 are discarded. For the remaining models, we choose the one which highest value in the main diagonal.

D. Interpretation of the Results

We evaluate the model with several scenarios with samples of actors of different sizes. We build a scenario C1 with a sample of size 350 and a second scenario C2 where the sample's size is 51. For each actor, we get the amount of his participation of the activities. We calibrate the mixture modeler and discover two GMM with $K=3$ and $K=2$ subgroups respectively for C1 and C2. The size of each cluster is resumed in table I. If a manager want to retrieve knowledge about how a given population should be populated into different distinct classes, he might observe and analyze only the results from the GMM clustering. If, instead, he want to go beyond this consideration and to learn more about the potential correlation between these families, he can follow this previous step by the HMM model calibration. So, after running this final operation, we found 2 super-groups (HMM states) for each scenario with the corresponding transition matrices:

$$A_{C1} = \begin{bmatrix} 0.9695 & 0.0305 \\ 0.0719 & 0.9281 \end{bmatrix},$$

$$A_{C2} = \begin{bmatrix} 0.9592 & 0.0408 \\ 0.0259 & 0.9741 \end{bmatrix}.$$

and corresponding emission probabilities:

$$B_{C1} = \begin{bmatrix} 0.0034 & 0.9816 & 0.0050 \\ 0.7540 & 0.0028 & 0.2433 \end{bmatrix},$$

$$B_{C2} = \begin{bmatrix} 0.9939 & 0.0061 \\ 0.0000 & 1.0000 \end{bmatrix}.$$

By observing the off-diagonal elements of the transition matrices, we can see clearly that there's a low probability to reach a state (or super-group) i from another different state j . The probability to reach a state being inside itself is high. In scenario C1, there's about 97% of probability of observing the state #1 and the probability to observe the state #2 is equal to 93%. For scenario C2, the states #1 and #2 are well-separated with probabilities 96% and 97% respectively. So we can conclude that the different super-groups are well-formed, clearly separated and dense.

Another interesting findings with the HMM model is related to the matrix B of emission probabilities. This matrix tells us the probability if observing the subgroups from each HMM state/super-group. So for scenario C1, see figure Figure 2, the first super-group is almost composed only by the subgroup #2 (with a probability of 98% of observing this cluster), where the second super-group is occupied both by the subgroup #1 with a probability of 75% and the subgroup #3 with 25% of emission probability. On the other hand, in scenario C2 where results are plotted in figure Figure 3, we have 2 subgroups which can be re-organized into two super-groups, where the first one is composed only with the subgroups #1 while the second super-group recover the subgroup #2. With this second scenario, the HMM/viterbi decoder does not discover a new reconfiguration of the subgroups. In figures 4 and 5, we plot the actors versus the amount of activities and show how each super-group capture the evolution and the spread of the data inside the different subgroups.

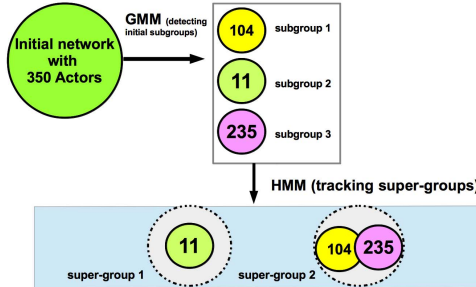


Fig. 2. Results corresponding to the detection of subgroups by the GMM and the tracking of super-groups by the HMM framework. The initial network is composed with 350 actors which participate to a huge number of activities. Case Scenario 1.

V. CONCLUSION

In this work, we have developed a robust framework based on parametric distributions able to solve two important problems in social network analysis. A social community is generally composed with individuals who may evolve with

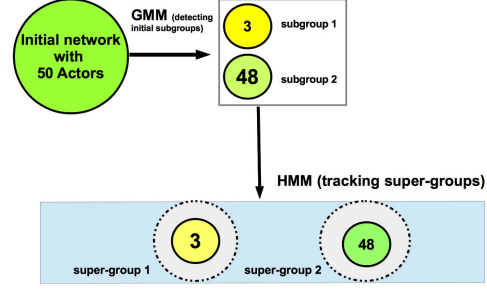


Fig. 3. Results corresponding to the detection of subgroups by the GMM and the tracking of super-groups by the HMM framework. The initial network is composed with 51 actors which participate to a huge number of activities. Case Scenario 2.

TABLE I
SIZE OF EACH SUBGROUP (GAUSSIAN COMPONENT).

Scenario C1			
subgroup i	1	2	3
Length subgroup i	104	11	235
Viterbi path			
super-group #1		super-group #2	
11		339	
Scenario C2			
subgroup i	1		2
Length subgroup i	3		48
Viterbi path			
super-group#1		super-group #2	
3		48	

both common and heterogeneous opinions and behaviors. In this work, we have done the assumption that these opinions/behaviors may come from a family of gaussian components. Consequently, calibrating a Gaussian Mixture Model can lead us to discover the different subgroups from an initial evolving community of actors and thus to solve the first problem. On the other hand, the second problem related to identifying potential dependencies between subgroups were decoded by means of a hidden Markov model. The result is that subgroups with self-similarities were merge to build new group. The great advantage to encounter these two problems is to facilitate decisions making for the managers and operators who put their business into social network behavior analysis.

In this work, there's an interesting remark about the use of a hidden Markov model to discover super-groups. Indeed, the observation (emission) probability which give the rate or membership degree of a given subgroup is not obviously equal to 1. This fact really signifies that a subgroup have already, for each super-group, a degree of membership such as a sort of probability to belong to a super-group. In this work, we have done a strong assumption making a given subgroup lying entirely to a specific super-group when its emission probabilities for the remaining super-groups are very law (<1%). At this moment, we can ask the following question. Is it possible to set a natural way to retrieve an automatic level or confidence bounds for the emission probability to

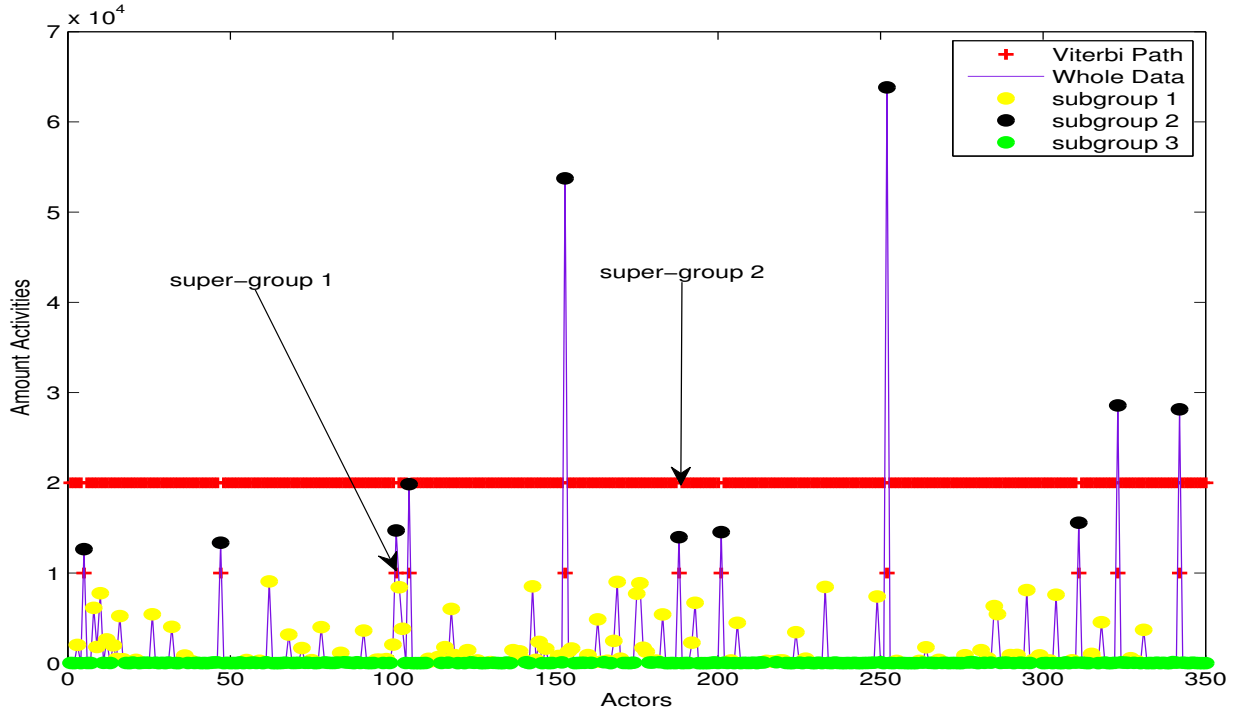


Fig. 4. Sub-community tracking. The Viterbi algorithm find the single best path trough the HMM trellis. The graphs show the whole actors and the activities, data corresponding to each cluster (subgroup) and the Viterbi path. The super-group 1 is composed only with subgroup 2 while the super-group 2 cover the subgroups 1 and 3. Case scenario 1.

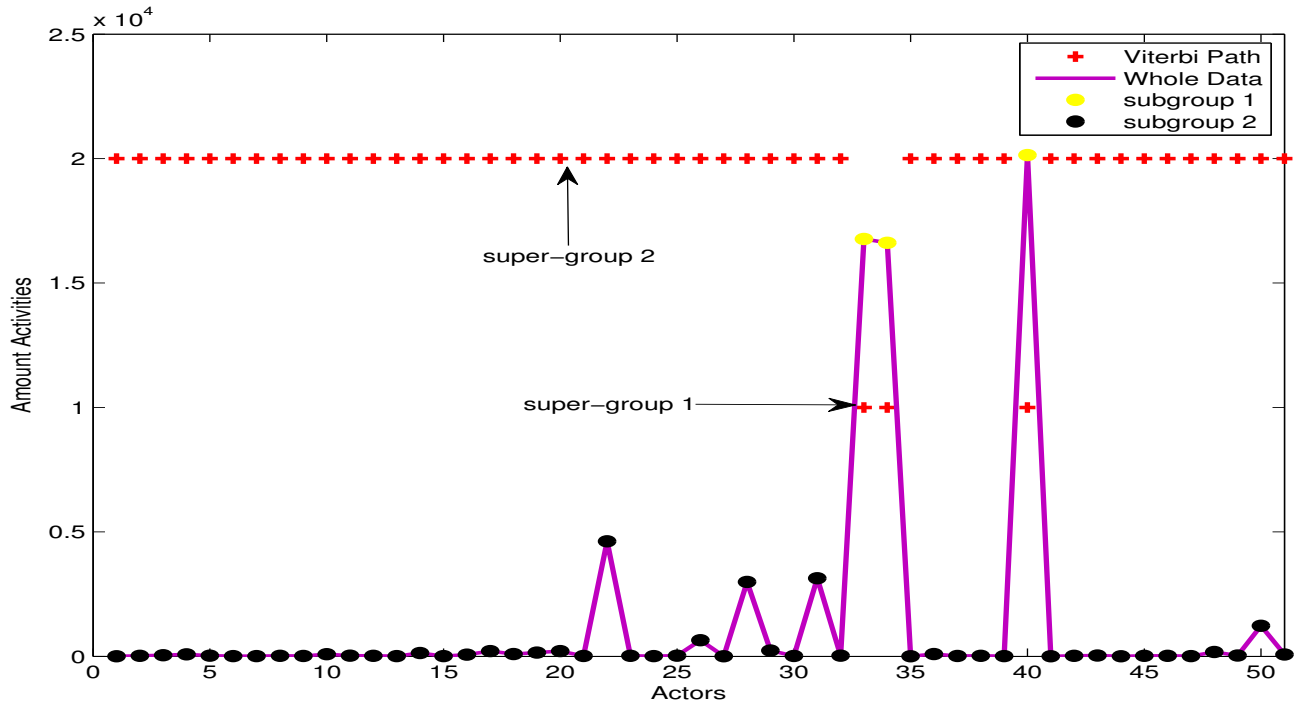


Fig. 5. Sub-community tracking. The Viterbi algorithm find the single best path trough the HMM trellis. The graphs show the whole actors and the activities, data corresponding to each cluster (subgroup) and the Viterbi path. The super-group 1 contains subgroup 1 while the super-group 2 cover the subgroup 2. Case scenario 2.

decide where a subgroup lies? In this situation, we can easily imagine that situations may happen where a given subgroup has rates which allow it to be member of several super-groups at the same time. Consequently, super-groups would be merged to form new super-group. We think also that the problems solved in this work may arise other questions about possible relationships over super-groups coming from different geographical area, making sense to talk about spatial correlations. We plan to overcome these questions and ideas in our future works.

REFERENCES

- [1] W. Avison, J. McLeod, and B. Pescosolido. *Mental Health, Social Mirror*. Springer, 2007.
- [2] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, pages 44–54, 2006.
- [3] B. Ball, B. Karrer, and M. E. J. Newman. An efficient and principled method for detecting communities in networks. *CoRR*, abs/1104.3590, 2011.
- [4] A. V. Carron and L. R. Brawley. Cohesion: Conceptual and measurement issues. *Small Group Research*, 31(1):89–106, 2000.
- [5] F. Chiocchio and H. Essiembre. Cohesion and performance: A meta-analytic review of disparities between project teams, production teams, and service teams. *Small Group Research*, 40(4):382–420, 2009.
- [6] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, February 2010.
- [7] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–7826, 2002.
- [8] M. K. Goldberg, M. Magdon-Ismail, and J. Thompson. Identifying long lived social communities using structural properties. In *ASONAM*, pages 647–653, 2012.
- [9] S. Kashoob and J. Caverlee. Temporal dynamics of communities in social bookmarking systems. *Social Network Analysis and Mining*, 2(4):387–404, 2012.
- [10] H. Lakkaraju, J. J. McAuley, and J. Leskovec. What's in a Name? Understanding the Interplay between Titles, Content, and Communities in Social Media. In *ICWSM*, 2013.
- [11] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1), Mar. 2007.
- [12] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.
- [13] J. Ndong. A new approach to anomaly detection based on possibility distributions. *INTERNET 2014, IARIA*.
- [14] J. Ndong. Using sub-optimal kalman filtering for anomaly detection in networks. *Proceeding of the Electrical Engineering Computer Science and Informatics*, 1(1):408–414, 2014.
- [15] J. Ndong and K. Salamatian. A robust anomaly detection technique using combined statistical methods. In *Communication Networks and Services Research Conference (CNSR), 2011 Ninth Annual*, pages 101–108. IEEE, 2011.
- [16] M. E. J. Newman. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2):321–330, Mar. 2004.
- [17] G. Palla, A.-L. Barabasi, and T. Vicsek. Quantifying social group evolution. *Nature*, 446:664–667, 2007.
- [18] T. Prohaska, L. Anderson, and R. Binstock. *Public Health for an Aging Society*. Public Health for an Aging Society. Johns Hopkins University Press, 2012.
- [19] R. Toivonen and L. Kovanen and M. Kiveli; and J. P. Onnela, and J. Sarami;ki and K. Kaski. A comparative study of social network models: Network evolution models and nodal attribute models. *Social Networks*, 31(4):240–254, October 2009.
- [20] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [21] I. Sarr and R. Missaoui. *Temporal Analysis on Static and Dynamic Social Networks Topologies*. Springer. To appear.
- [22] I. Sarr, J. Ndong, and R. Missaoui. Overlaying social networks of different perspectives for inter-network community evolution. In *Social Network Analysis-Community Detection and Evolution*, pages 45–70. Springer, 2014.
- [23] K. M. Shapcott, A. V. Carron, S. M. Burke, M. H. Bradshaw, and P. A. Estabrooks. Member diversity and cohesion and performance in walking groups. *Small Group Research*, 37(6):701–720, 2006.
- [24] J. Yang, J. McAuley, and J. Leskovec. Detecting cohesive and 2-mode communities in directed and undirected networks. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 323–332, New York, NY, USA, 2014. ACM.
- [25] S. Zhang, H. Jiang, and J. Carroll. Integrating online and offline community through facebook. In *Collaboration Technologies and Systems (CTS), 2011 International Conference on*, pages 569–578, May 2011.