COGNISON: A Novel Dynamic Community Detection Algorithm in Social Network

Hamideh Sadat Cheraghchi*
Department of Computer Engineering and Science, Shahid Beheshti University, Tehran, Iran
h.s.cheraghchi@gmail.com
Ali Zakerolhossieni
Department of Computer Engineering and Science, Shahid Beheshti University, Tehran, Iran
a-zaker@sbu.ac.ir

Received: 06/Mar/2016 Revised: 15/May/2016 Accepted: 25/May/2016

Abstract

The problem of community detection has a long tradition in data mining area and has many challenging facet, especially when it comes to community detection in time-varying context. While recent studies argue the usability of social science disciplines for modern social network analysis, we present a novel dynamic community detection algorithm called COGNISON inspired mainly by social theories. To be specific, we take inspiration from prototype theory and cognitive consistency theory to recognize the best community for each member by formulating community detection algorithm by human analogy disciplines. COGNISON is placed in representative based algorithm category and hints to further fortify the pure mathematical approach to community detection with stabilized social science disciplines. The proposed model is able to determine the proper number of communities by high accuracy in both weighted and binary networks. Comparison with the state of art algorithms proposed for dynamic community discovery in real datasets shows higher performance of this method in different measures of Accuracy, NMI, and Entropy for detecting communities over times. Finally our approach motivates the application of human inspired models in dynamic community detection context and suggest the fruitfulness of the connection of community detection field and social science theories to each other.

Keywords: Social Network; Clustering; Cognitive Modeling; Evolution.

1. Introduction

The twist from self-reported survey data to autonomous data gathering enabled by Web 2 and new technologies e.g. smart phones, email, and other smart data gathering gadgets change the dimension and order of data to be analyzed unprecedentedly. Mining such data to recognize and track linked interactions of individuals is a critical area of interest for analyst due to its wide application from cybersecurity to recommender and trade systems. This problem known as community detection problem in social network is one of the well-studied areas of research during the past decades and is linked to general data clustering problem. The existence of linked data is distinguishing feature of modern community detection in social network context versus traditional point-based data clustering.

Various challenging facets of community detection has brought different solutions to this problem from computer engineering, physics, and social science perspectives and changed it to a multi-disciplinary problem. The well-studied statistical inference-based models [1], hierarchical algorithms [2], spectral and modularity-based models [3] are among these models. Survey papers [1-3] are referred for a complete review. These techniques are designed basically to capture the communities in static networks; i.e. network data is gathered in one time step or in the case of multiple time steps, data is mixed to have the picture of the network in one snaps. With rapid growth of online social networks, where users' joining in and withdrawing from communities are common, dynamic network evolution is recognized as a very valuable research domain for content management and recommender systems [4]. Designing such dynamic algorithms to capture different events happening in the network has its own challenges. Accounting for unforeseen change in topological structure and at the same time temporal smooth overlapping structure are among these challenges. Meanwhile, most algorithms in dynamic settings are extension of static algorithms which will be discussed in Section 2.

Interestingly, social scientists were the prime group of researchers who were always attracted to study the whole network evolution and dynamism of individuals' interaction over different time steps to find its underlying principles. There are numerous school of thoughts for exploring the principles behinds individual mechanisms leadings to versatile network dynamism and community evolution. The famous sociologist, Barry Wellman [8] introduced five main principles to explore intellectual disciplines underlying networks. These fundamental principles mainly focus on relation of individuals to each other rather than individual attitude or demographic characteristic for predicting their behavior. He emphasizes on the dynamic context of relationships and the imposed effects of relationships on everyone in the network. Further,

* Corresponding Author

there are numerous theoretical roots on why people create and maintain groups as discussed in [9] including theories of self-interest, social exchange or dependency, mutual or collective interest, homophily and cognitive theory. Prototype theory as a model for selecting the groups to join is also derived from cognitive science findings.

In this research, we propose our community detection computational model by taking advantage of related social theories devised for exploration of dynamic behavior of human in joining/leaving communities in social network. For this purpose, we base our community detection algorithm according to general approach that human uses for selection: i.e. selection according to prototypes. Then, tracking the evolution of communities is achieved by following the dynamic relationship between entities and the communities and among the communities themselves. Hence, our algorithm is placed in category of prototype-based algorithms like k-means. Prototype based algorithms needs similarity measure for assigning nodes to communities. For this, cognitive consistency theory is our choice among various cognitive theories proposed. This theory helps to explain the natural tendency of human to decrease conflicting cognitions and attach to groups whom feels similar. This is also explained in homophily theory which explains willingness of individuals to communities where there are peoples one deems to be similar. In fact, similarity helps to reduce potential conflicts and increase predictability of behaviors. These two theories are related since choosing similar other persons reduces possible conflicts and increase consistency among members. Leveraging these disciplines, we devise our algorithm called COGNISON (COGnitive inspired community detection in Social Network) to tackle community detection problem in dynamic setting without any parameter or initial settings. Besides, our algorithms works for both weighted and binary networks which make it a preferable choice for social community detection problems.

The paper is organized as follows: Section 2 presents a preview of common background knowledge in community detection context. Section 3 explains our proposed approach and Section 4 represents the experiments to evaluate COGNISON. Finally, we conclude with some highlights on our future research directions.

2. Related Work

As already discussed, traditional community detection approach are designed intrinsically to capture the communities in static networks which limits the analysis of the networks and ignores events which may be much easier to predict if one had the whole picture of the network in different time steps. Hence, a new line of research has been developed focusing on tracking communities and events happening during different time steps; i.e. dynamic dimension of communities [4]. Here, we give a summary of two main lines of research for studying the evolution of communities.

The first and the most intuitive approach uses some static community detection algorithms in each time step of the network. The changes undergone by the communities in different time steps are tracked according to some similarity/distance measures such as intersection of communities to determine the relationship among communities thereby tracking dynamicity of communities. This approach is called *independent community mining*. The main characteristic of this approach is discovering communities from the scratch in each time step using independent or equal algorithms. In the other category, there are algorithms that incorporate the information obtained in other snapshots for extracting communities of future time step and is called *incremental community detection*. This approach of algorithm improves the time and computational complexity compared to independent community detection approach [5-7]. In one dominant approach in this category, a cost-function is calculated in each time step trying to minimize the changes happening to communities in the following time step. This approach assumes that abrupt changes in subsequent time steps are unlikely and these changes have small impact on the community structure. This concept was introduced by Chakarbarti et al.[8] who coined the term *evolutionary clustering* in which two potentially contradicting criteria in an additive equation should be optimized. The first criterion is the correspondence of clustering result to current data as much as possible (*clustering quality*) and the second is keeping the shifts of the results between current clustering and previous time step as low as possible (*history quality*) to allow for *temporal smoothness* as formulated in equation 1. Notice that cost-based approaches is unable to handle drastic changes happening in the network In fact, the assumption of small changes in the network in cost-based approach limits its applicability when abrupt changes happens due to different reasons. Problems with choosing the ideal value for smoothing parameter (α in equation 1) responsible for tuning the weight to place on historic or clustering quality is also addressed in [9].

$$\mathrm{Cost_{total}} = \alpha \mathrm{Cost_{quality}} + (1-\alpha)\mathrm{Cost_{quality}} \qquad (1)$$

Furthermore, there are some other *direct methods* with different definitions for quality and temporal cost. Whatever the definition is, cost functions are incorporated in different static clustering modularity [10], spectral [11], and inference-based [12] paradigms to capture the dynamic of the network. Survey papers [4,13,14] are used for a complete review. Following we review shortly some other important group of community detection paradigms.

In *Modularity-based algorithms* dense communities are recognized using modularity criterion [15,16] in which agglomerative algorithms greedily optimize modularity criterion. Modification of these algorithms is the basis of dynamic modularity based algorithms. For example, Gorke et al. [10] take the changing nodes in different time steps into a separate community and revise the membership according to some merge functions based on modularity criterion. However, modularity-based community detection algorithms bear some weaknesses [17]: inability to handle noise and a large number of high score communities which avoid recognition of specific community structures [18] are among these problems. *Evolutionary spectral clustering* approach finds an n-dimensional placement of nodes according to a variation of adjacency matrix, e.g. eigenvalues as a cost function to regularize temporal smoothness [11] or investigate the changes to eigenvalues in different time steps [19]. The weakness of spectral-based clustering algorithms lies in high computational cost incurred during matrix multiplication which makes it a weak choice for very large networks. *Inference-based methods* are also another broad category which considers an underlying statistical model that can generate communities in the network. FacetNet [12] is the first probabilistic generative model proposed for analyzing evolution of communities and several other works have been introduced recently [20]. This category also suffers from high memory usage [21].

Usage of cognitive and social theories in clustering domain is also taking new dimensions recently. Apart from famous k-means algorithm and its derivations for example belief k-mode clustering [22] which are all placed in this category, cognitive inspired clustering are leveraged in different applications. Data dissemination based on cognitive theories [23] and linguistic clustering by prototype theory [24] are among recent approaches in this category. Human group formation based on homophily or similarity among members is also verified in different studies [25]. Now, we take advantage of social theories to introduce a new community detection algorithm in social networks applicable in both binary and weighted networks. In contrast to k-means like algorithms, the number of communities will be determined automatically according to characteristic of the network.

3. Proposed Model

Suppose a network G with n node where the interaction between each pair of members in time *t* is indicated by a symmetric binary/weighted link. This information can be represented in a proximity matrix W^t where W_{ij}^t denotes the proximity between member *i* and *j* according to the link weight of connected items (edge weight of binary network is 1).

Now, we describe how cognitive inspired disciplines helps to design an efficient community detection algorithm for tracking the evolution of communities. For this purpose, we leverage the approach that human uses for categorization: i.e. prototype based selection in our community detection algorithm and track the evolution of communities by following the dynamically updated structures between entities and the communities in the history of the network. In fact, communities are created and altered online in regard to the observed changes in the network. How to follow this relationship is the key to track the dynamic of communities. We explore functionality of this algorithm in two key phases of recognition and categorization. In the *recognition* phase, members are introduced to communities present in the networks and selection process according to a cognitive inspired similarity measure happens. In the second phase of *learning*, prototypes are updated to reflect the events happened in the network and set as proper candidate for future selection scenarios. These steps are explained in details in the next sections.

3.1 Recognition Phase

In this phase, members explore the environment and a selection process takes place similar to operation of prototype theory. Each input entry *i* along with its interacting neighbors recorded in W_i^t finds community prototypes available from past history and decides on the best one to join. This is through similarity checking of objects against prototypes. Now, we elaborate on the details of structures and similarity measure used in COGNISON.

Input entry leverages the minimum local information available of its own id and their neighbor ids, the frequency of interaction with its neighbor and the time of interaction to construct its own input structure and prototype structure. Hence, minimal features stored in prototype structure are ID member, frequency and time step of linked data observed (<*id*,*frequency*,*time*>). Obviously, at the beginning of the first time step, there is no community prototypes and the values of the first linked input entry are stored in the first community prototype. As the subsequent entries are entered, they are checked by a similarity measure to see whether they can be included in one of the existing prototypes or a new one should be created to accommodate properties of this entry. This process continues until all nodes in current time step are assigned to their communities. At beginning of next time steps, the available prototypes derived in previous time step are leveraged for comparison purpose. This is compatible with the idea that members tend to preserve their membership to their assigned communities in previous time steps.

Now, we elaborate on how to compute similarity measure of each data entry to available category prototypes. Assignment to one of the prototypes is steered by cognitive consistency theory. For this, we use finding of experiments directed by behavioral researchers in which to assess predictability of friends' revisit in the future, one should consider the effects of *frequency* and *spacing* pattern of previous visits as major elements for prediction of future visits [26,27]. In their definition, frequency is the rate of previous visits and recency is the time elapsed since last visits. Whenever frequency and recency of visit is high, there is high possibility of revisit. We use this concept to categorize each input entries in the most similar community prototype and derive our measure to assess similarity of input entry I to each prototype $C_i^* \in C^*$ as follows:

$$Likeness(I, C_k^*) = \sum_{j \in M} Activity_J * Recency_J \quad (2)$$

$$\mathrm{Recency_J} = \mathrm{e}^{\frac{\mathrm{t-t_j}}{\mathrm{t}}} \quad (3)$$

In this equation, similarity is computed in *Likeness* formula which is computed based on two features of common members between input entry I and examined prototype C_i^* (M denotes common members). The more common members exists in one prototype, the higher is the chance of selecting it. However, *activity* and *recency* of those common members in each prototype are other important factors for selection. *Activity* of a member in prototypes is the weights of its interaction when it is included in the prototype. Hence, more a node is observed, its activeness will be higher. For weighted networks, each observation of input entry will record the weight of interaction. Further, *recency* of each member is computed using an exponential function which takes into account the difference among current time and the last time the entry the member is observed. So, if the entry is observed in community in just current time step, this variable takes its highest value of 1 and if observed in older time steps, it acquires a value less than 1 which will decrease the previously seen activity of that member. If more than one community takes non-zero value of similarity, the one with highest measure is selected for inclusion of input entry.

3.2 Learning Phase

After the assignment of nodes to its best recognized communities, an update scheme should take place in the selected prototype to reflect the changes made due to recently added member structure assignment. If a member belongs to more than one community in the final stage, we assign the node to the community in which the node has the highest similarity. In the update process, three main feature of prototypes, i.e. <*id,frequency,time*> each are updated. ID of new member is added if not already present in the prototype structure and *frequency* of interaction is updated by summing up current frequencies of input entries to their old values present in the prototype. Finally, time property of the members present in the current time step is updated. In this way, activity of nodes which have not been observed in previous time step is decreased which helps the algorithms to be responsive to new events while preserving past events. This is achieved when exploiting *likeness* measure (eq. (2) and (3)) for selecting the best prototype.

4. Experimental Results

We examine the performance of the proposed algorithm on both evolving synthetic and real datasets. In the synthetic experimental section, in addition of toy example, we use the stabilized and frequently used synthetic LFR generator is used for artificial dataset generation and for real dataset experiment, the famous of MIT really mining dataset is exploited. The number of entities in synthetic and real dataset may change in different time steps. Further, the numbers of communities differs in the intervals. Since ARTISON inherits most of its properties from representative-based algorithms, proper comparison is achieved by comparing it to other representative-based algorithms. For this reason, we choose two state-of-art evolutionary k-means algorithms specially designed for dynamic settings of network for comparison purpose. This equals to compare ARTISON with the pioneer evolutionary framework extended to k-means [28] where two temporal and quality costs are optimized with a constant smoothing factor (α in equation (1)) to capture the dynamic of the network. Further, we use another more recent evolutionary framework extended to k-means algorithm called Adaptive Evolutionary Clustering (AFFECT [29]) where optimal smoothing factor is determined automatically using a statistical approach. In all of these case, the optimum number of communities for each time step is determined by well-known silhouette width criterion [30]. This measure determines how compact the distance of communities are in a given time step and the maximum width of this measure is used to assess the number of needed communities in k-means. In addition, we use two other modern hierarchical agglomerative community detection algorithms based on modularity criterion in social network for real dataset experiments to make comparison with state of art algorithms. Louvain [31] and fast modularity [32] where both have acquired high performance in recent survey studies.

For the evaluation, we use four measures to determine the accuracy and quality of the community detection algorithms in different time steps via clustering Rand Index and F measure to indicates the amount of disagreement between discovered communities (C) and the labels of ground truth communities (C^*). F measure is a harmonic mean of precision and recall measures where precision is the ratio of relevant objects (real community member detected) to total number of objects detected and recall is the ratio of relevant objects detected to total real ground truth members. All the mentioned measures reach their best at 1 and their worse at 0 value.

$$\mathrm{F(C, C^*)} = \frac{2|\mathrm{C} \cap \mathrm{C}^*|}{|\mathrm{C}| + |\mathrm{C}^*|} \quad (4)$$

Higher values of all of these measure are preferred. For quality of clustering we use another common measure in information theory called Entropy [33] to measure the quantity of the disorder observed in the results. Lower value of entropy is preferred which means better clustering result.

4.1 Synthetic Dataset Evaluation

In the first experiment, we use a toy synthetic dataset in which different numbers of communities appear during the test to better verify the dynamic community tracking capabilities of the proposed algorithm. Figure 1 shows the diagram of the synthetic dataset.

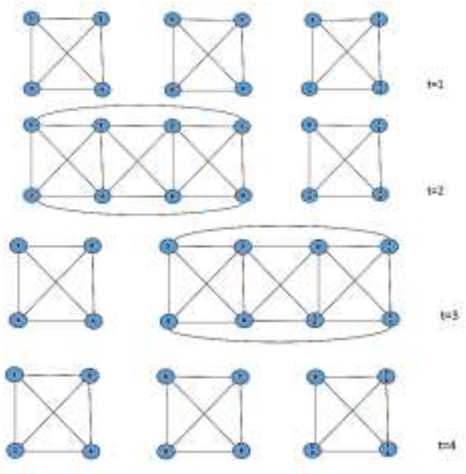

Fig. 1. An example of community evolution in four time step: in t=1, three communities, in t=2 and t=3 two communities and t=4 again three community exists.

As indicated in the figure, in the first time step, three communities are recognizable. Then, a merging happens and two communities are distinguishable in t=2 and t=3. Finally, we come back again to three communities observed first. For LFR generator, we use two experiments of switch and expand/contract events to test the performance of the algorithm. The other parameters of the generator are set as follows: average and maximum degree of node is set to 10 and 50, and minimum and maximum community size is 20 and 100 nodes respectively and the initial number of nodes is 500 nodes. The experiments are averaged over different runs since k-means based algorithms produce different results in multiple run due to initial node. In the switching event, the number of nodes during the whole experiment is fixed but they change their communities with probability of 20%. For the expansion event, we considered three expansions and two contraction events per time step by switching probability of 10% (50 nodes out of 500 nodes switch their community in each time step).

Figure 2 shows the result of competing algorithms in terms of Rand index and F measure. We presented the mean and standard error over all time steps. Since we expect a higher value for Rand index and F-measure for recognition of the preferred community detection approach, we judge COGNISON as the superior one. For the standard deviation, COGNISON in several cases has higher deviation from other algorithms but high difference of the measure compared justify this variation.

Table 1. Comparison of proposed algorithm in synthetic datasets with other protocols in two measures of a) Rand Index, b) F measure.

Dataset	Measures	COGNISON	AFFECT	Evolutionary-k-means
Toy Dataset	Rand	0.86±0.16	0.70±0.05	0.72±0.15
	F	0.80±0.25	0.58±0.14	0.60±0.26
Switch (0.2)	Rand	0.82±0.05	0.42±0.07	0.44±0.10
	F	0.39±0.19	0.21±0.10	0.31±0.13
Expand	Rand	0.86±0.06	0.36±0.15	0.39±0.14
	F	0.51±0.23	0.22±0.14	0.26±0.14

4.2 Real Dataset Evaluation

In this section, we intend to evaluate the performance of COGNISON on the Reality Mining dataset [34] commonly used for dynamic evaluation purpose [35,36]. This dataset is gathered by MIT media lab to analyze the cell phone activity of 90 participants consisting of students and staff interacting over a period of nine months. The large volume of approximately 500,000 hours of data is extracted by monitoring different cell usage of participants logged as incoming and outgoing calls, cell tower id, and any Bluetooth devices discovered during their interactions. Our experiments covers Bluetooth activity of participants which records the IDs of nearby Bluetooth devices (student or student ID) every five minutes. Affiliation of each participant is available to be used as the ground truth information as discovered by Eagle et al. [34]. Further, for finding optimal number of communities, we use silhouette measure [30].

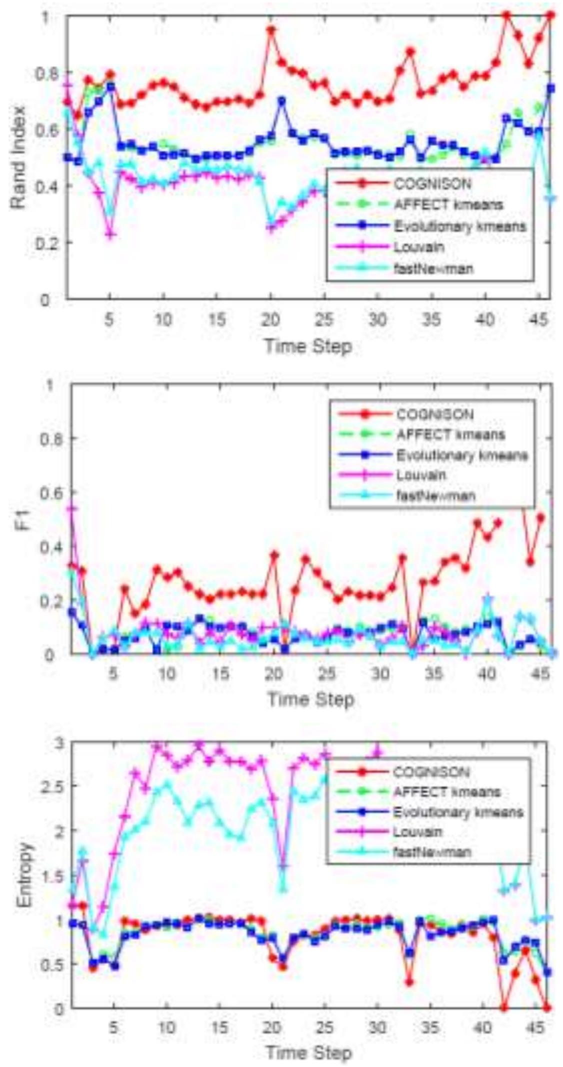

Fig. 2. Comparison of Rand Index and F measure values achieved in real dataset for the five algorithms: COGNISON, AFFECT, evolutionary k-means, Louvain and Fast Newman clustering algorithms.

Table 2. Comparison of proposed algorithm in synthetic datasets with other protocols in two measures of a) Rand Index, b) F measure.

	Rand Index	F-Measure	Entropy
COGNISON	0.75±0.06	0.26±0.07	0.76±0.27
AFFECT	0.55±0.07	0.07±0.04	0.84±0.17
Evolutionary k-means	0.55±0.07	0.07±0.04	0.83±0.15
Louvain	0.41±0.09	0.08±0.08	2.30±0.60
Fast Newman	0.43±0.08	0.07±0.06	1.98±0.45

For entropy measure, lower value shows less quantity of disorder found in community detection and is desired. As depicted in the last row of Table 2, entropy of COGNISON is lower than all.

Notice that COGNISON has several distinguishing features. The ability of discovering the number of communities intrinsically is of great advantage while the other k-means based algorithms assess the number of optimal characters as inputs or find it through some other calculations external to the algorithm (e.g. using silhouette [37] as used in our experiments) for the initialization of the algorithm. Second, the algorithm is free of any smoothing factor commonly used for evolutionary algorithms. In fact, in COGNISON, there is no tradeoff between quality and history costs which makes it more robust to changes.

5. Conclusion and Future Works

Following the stream of works presented for dynamic community detection, specifically evolutionary clustering algorithms, we proposed another online dynamic community detection algorithm in social network context called COGNISON. While the initialization of each time step takes community snapshot of the previous time step into account, different mechanisms for link weighting cause the enforcement of strong links and weakening of weak links not present in the previous steps. This helps to solve a big challenge in most community detection algorithms; i.e. knowing the number of communities to pass as an input to the algorithm. The experimental results displayed the good performance of this algorithm against the state of art evolutionary algorithms and encourage the ongoing works on dynamic community detection to take more advantage of cognitive-inspired paradigms.

References

[1] D. J. MacKay, Information theory, inference and learning algorithms: Cambridge university press, 2003.

[2] K. Sasirekha and P. Baby, "Agglomerative hierarchical clustering algorithm–A Review," International Journal of Scientific and Research Publications, vol. 3, 2013.

[3] T. N. Dinh and M. T. Thai, "Community detection in scale-free networks: approximation algorithms for maximizing modularity," Selected Areas in Communications, IEEE Journal on, vol. 31, pp. 997-1006, 2013.

[4] T. Aynaud, E. Fleury, J.-L. Guillaume, and Q. Wang, "Communities in evolving networks: Definitions, detection, and analysis techniques," in Dynamics On and Of Complex Networks, Volume 2, ed: Springer, 2013, pp. 159-200.

[5] M. Takaffoli, J. Fagnan, F. Sangi, and O. R. Zaïane, "Tracking changes in dynamic information networks," in Computational Aspects of Social Networks (CASoN), 2011 International Conference on, 2011, pp. 94-101.

[6] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks," in Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on, 2010, pp. 176-183.

[7] T. Falkowski, J. Bartelheimer, and M. Spiliopoulou, "Mining and visualizing the evolution of subgroups in social networks," in Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 2006, pp. 52-58.

[8] D. Chakrabarti, R. Kumar, and A. Tomkins, "Evolutionary clustering," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 554-560.

[9] J. Zhang, Y. Song, G. Chen, and C. Zhang, "On-line Evolutionary Exponential Family Mixture," in IJCAI, 2009, pp. 1610-1615.

[10] R. Görke, P. Maillard, A. Schumm, C. Staudt, and D. Wagner, "Dynamic graph clustering combining modularity and smoothness," Journal of Experimental Algorithmics (JEA), vol. 18, p. 1.5, 2013.

[11] Y. Chi, X. Song, D. Zhou, K. Hino, and B. L. Tseng, "Evolutionary spectral clustering by incorporating temporal smoothness," in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 2007, pp. 153-162.

[12] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, "Facetnet: a framework for analyzing communities and their evolutions in dynamic networks," in Proceedings of the 17th international conference on World Wide Web, 2008, pp. 685-694.

[13] S. Papadopoulos, Y. Kompatsiaris, A. Vakali, and P. Spyridonos, "Community detection in social media," Data Mining and Knowledge Discovery, vol. 24, pp. 515-554, 2012.

[14] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, "Analyzing communities and their evolutions in dynamic social networks," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 3, p. 8, 2009.

[15] M. E. Newman, "Fast algorithm for detecting community structure in networks," Physical review E, vol. 69, p. 066133, 2004.

[16] S. Cafieri, P. Hansen, and L. Liberti, "Locally optimal heuristic for modularity maximization of networks," Physical Review E, vol. 83, p. 056105, 2011.

[17] E. Eaton and R. Mansbach, "A Spin-Glass Model for Semi-Supervised Community Detection," in AAAI, 2012.

[18] B. H. Good, Y.-A. de Montjoye, and A. Clauset, "Performance of modularity maximization in practical contexts," Physical Review E, vol. 81, p. 046106, 2010.

[19] H. Ning, W. Xu, Y. Chi, Y. Gong, and T. S. Huang, "Incremental spectral clustering by efficiently updating the eigen-system," Pattern Recognition, vol. 43, pp. 113-127, 2010.

[20] L. Tang, H. Liu, and J. Zhang, "Identifying evolving groups in dynamic multimode networks," Knowledge and Data Engineering, IEEE Transactions on, vol. 24, pp. 72-85, 2012.

[21] M. B. Hastings, "Community detection as an inference problem," arXiv preprint cond-mat/0604429, 2006.

[22] S. B. Hariz, Z. Elouedi, and K. Mellouli, "Clustering approach using belief function theory," in Artificial Intelligence: Methodology, Systems, and Applications, ed: Springer, 2006, pp. 162-171.

[23] M. Conti, M. Mordacchini, and A. Passarella, "Design and performance evaluation of data dissemination systems for opportunistic networks based on cognitive heuristics," ACM Transactions on Autonomous and Adaptive Systems (TAAS), vol. 8, p. 12, 2013.

[24] H. Zhao and Z. Qin, "Clustering Data and Vague Concepts Using Prototype Theory Interpreted Label Semantics," in Integrated Uncertainty in Knowledge Modelling and Decision Making, ed: Springer, 2015, pp. 236-246.

[25] N. F. Johnson, C. Xu, Z. Zhao, N. Ducheneaut, N. Yee, G. Tita, et al., "Human group formation in online guilds and offline gangs driven by a common team dynamic," Physical Review E, vol. 79, p. 066117, 2009.

[26] M. M. Krasnow, A. W. Delton, J. Tooby, and L. Cosmides, "Meeting now suggests we will meet again: Implications for debates on the evolution of cooperation," Scientific reports, vol. 3, 2013.

[27] T. Pachur, L. J. Schooler, and J. R. Stevens, "We'll Meet Again: Revealing Distributional and Temporal Patterns of Social Contact," 2014.

[28] Y. Chi, X. Song, D. Zhou, K. Hino, and B. L. Tseng, "On evolutionary spectral clustering," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 3, p. 17, 2009.

[29] K. S. Xu, M. Kliger, and A. O. Hero Iii, "Adaptive evolutionary clustering," Data Mining and Knowledge Discovery, vol. 28, pp. 304-336, 2014.

[30] L. Kaufman and P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis vol. 344: John Wiley & Sons, 2009.

[31] 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, p. P10008, 2008.

[32] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," Physical review E, vol. 70, p. 066111, 2004.

[33] Y. Yao, "Information-theoretic measures for knowledge discovery and data mining," in Entropy Measures, Maximum Entropy Principle and Emerging Applications, ed: Springer, 2003, pp. 115-136.

[34] N. Eagle, A. S. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," Proceedings of the National Academy of Sciences, vol. 106, pp. 15274-15278, 2009.

[35] J. More and C. Lingam, "Current trends in reality mining," ed: IRJES, 2013.

[36] H. Zhang, R. Dantu, and J. W. Cangussu, "Socioscope: Human relationship and behavior analysis in social networks," Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, vol. 41, pp. 1122-1143, 2011.

[37] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," Journal of computational and applied mathematics, vol. 20, pp. 53-65, 1987.

Hamideh Sadat Cheraghchi received the BSc degree from Azad University, South Tehran branch, Tehran, Iran in 2006 MSc. from Azad University, Qazvin branch in 2009 and she is currently a PhD candidate in computer architecture in the department of Computer Engineering and Science at Shahid Beheshti University, Iran. Her research focuses on data mining in social network under the supervision of Dr. Ali Zakerolhosseini.

Ali Zakerolhoseini received the BSc degree from university of Coventry, UK, in 1985, MSc from the Bradford University, UK, in 1987, and PhD degree in Fast transforms from the University of Kent, UK, in 1998. He is currently an assistant professor in the department of Computer Engineering and Science at Shahid Beheshti University, Iran. His research focuses on Reconfigurable device and multi classifiers. His current research interests are Data Security, Cryptography and Social Data Mining.