Research Article

Implementation and Analysis of Clustering Techniques Applied on Pocket Switched Network

Muhammad Ali, 1 Mah-Rukh Fida, 2 Ameer Shakayb Arsalaan, 3 and Awais Adnan 1

1 Department of Computer Science, Institute of Management Sciences, Peshawar 25100, Pakistan
2 Shaheed Benazir Bhutto Women University, Peshawar 25000, Pakistan
3 COMSATS Institute of Information Technology Attock, Attock 43600, Pakistan

Correspondence should be addressed to Muhammad Ali; muhammad.ali@imsciences.edu.pk

Received 10 April 2015; Revised 13 August 2015; Accepted 1 September 2015

Academic Editor: Federico Barrero

Clusterings is an extraction of closely knitted groups from a set of nodes. Its benefits in social network range from applying marketing schemes on an appropriate interest group to social network analysis. It is also considered an important tool for efficient communication in an intermittent Pocket Switched Network (PSN). Contact probability between mobile devices in disrupted social networks greatly depends upon the mobility profile and level of relationships between the device holders. Unlike flat routing, scalable and efficient routing in these networks is highly dependent upon accurate derivation of social circles or clusters. This paper therefore evaluates existing clustering techniques for terrestrial social network with the end aim of minimizing communication overhead by identifying those message carriers that can bring message closer to destination node. In order to ensure intercluster routing, modification in existing schemes is proposed so as to detect bridge nodes between single hop destination clusters and to find path towards a disjoint destination cluster.

1. Introduction

Social network is made of people tied with each other due to common interest and sameness of geographic location and of work place. The advent of social media such as Twitter and LinkedIn has given an opportunity to like-minded and known people to build up virtual communities by creating, sharing, and exchanging mutually interesting information.

An important process applied on social network is clustering. It extracts closely related groups of nodes called clusters such that the relationship among nodes of two different clusters is nearly nonexistent. It is a beneficial tool for social media in the sense that without leaking out private information like names, addresses, and contact numbers of individuals, collaborating in a common field, it can track the group for a needy person to get help and support from.

This paper aims to reveal the effectiveness of existing clustering techniques, currently used in social networks, on social Pocket Switched Networks (PSN) [1]. PSN comprises of handheld mobile devices where there is no infrastructure like cellular technology and Internet and where owners have to depend upon bluetooth or Wi-Fi radios to connect with other persons.

PSN cannot use traditional routing protocols due to lack of continuous end-to-end connectivity compared to traditional networks. In PSN, communication is performed by opportunistic meeting of mobile devices’ owners. This meeting enables devices to come within each other’s range and exchange messages. In order to avoid the burden on resources and perform efficient transmission of message towards destination, the stable social relationship of the owners of the portable devices is exploited.

To forward messages only to those neighbors that can access destination device one has to identify nodes that are members of destination cluster. This demarcation of clusters helps in efficient and scalable routing as is done in CRHC [2] which is a hierarchical cluster-based scheme proposed in order to handle routing in a large, complex, and disrupted network. Similar is the case with ACHR (Adaptive Clustering Hierarchy Routing) [3] which is a hybrid hierarchical and cluster oriented scheme where nodes detect their clusters using local information with two-hop local visibility. In
this scheme, single-copy and multi-hop mechanism is used within a cluster whereas multi-hop and hop-by-hop strategy is used between clusters for message delivery.

Similarly, clustering is also utilized in cluster-based routing as shown in [4] where nodes visiting similar locations are grouped into a single cluster. The purpose is for co-cluster members to share resources, assist each other in load balancing, and reduce overhead so as to enable scalability and enhance efficiency in routing.

In this paper, we have applied some of the well-known and state-of-the-art clustering techniques used in the social networks domain. The effectiveness of these schemes is assessed by four different metrics, namely, modularity, complexity, conductance, and expansion (these four metrics are discussed in Section 4). Another aim of the paper is to view clustering as a means for proficient communication. Clustering algorithms, however, do not derive links between clusters while intelligent cross-cluster message transmission needs revealing of paths between source and destination clusters. We aim to supplement clustering techniques with uncovering of these paths between clusters, thus enabling routing protocols to be more efficient and resource saving.

The paper has been organized into six sections. Section 1 gives the introduction to the paper. Section 2 discusses state-of-the-art clustering techniques to be applied on social networks while Section 3 describes identification of bridge nodes and paths between pair of clusters. Section 4 presents the real world datasets of social network and quality functions on the basis of which efficiency of clustering technique is assessed. Section 5 interprets results of simulation runs while Section 6 concludes the paper with future work.

2. Literature Review

The role of clustering has been found to be significant for various domains including web content extraction [5], food webs, ecological networks [6], email networks [7], animal social networks, and gene networks [8]. Clustering techniques have also been used in ad-hoc wireless networks for scalability and energy efficiency [9]. Cluster-based routing in wireless networks increased route lifetime and decreased amount of control information to be stored and propagated in the network.

Clustering has been proposed for routing in social disrupted networks such as PSN. Global provision of Internet anytime and anywhere is not possible; therefore message transmission has to rely on local connectivity. Though this local connectivity can carry heavy message due to large bandwidth, yet due to limited radio ranges, the end points need to be within each other’s contact range. A method is therefore required which can perform message transfer even if end points are away from each other, that is, by using relays that might carry message closer to the destination. These relays are identified by using the stable knowledge of social clusters.

As communicating mobile devices are held by human owners, they have natural tendency to form clusters due to variation in relationships like friends, family members, familiar strangers, and strangers [10]. The basic aim is to identify these clusters which will ultimately make routing protocol scalable, more effective, and resource saving.

One of the important algorithms in the literature that can be applied on the networks mentioned above is given by Girvan and Newman in [11]. It is a hierarchical algorithm that initially considers the whole network as a single component and divides it into subcomponents by splitting the edge having highest betweenness. Betweenness of an edge is the number of times this edge lies on shortest paths between different pair of nodes. Costa et al. [12] have used Girvan and Newman algorithm for publish-subscribe applications that require dissemination of information to a group of interested people in an intermittent network.

K-Means [13], a simple clustering algorithm, was used by Jerusha et al. [14] to group sensor nodes on the basis of their geographic distances from K centroids and their energy strengths. These clusters were assigned cluster heads so as to avoid communication overhead among sensor nodes. Bahmik et al. [15] also used a type of K-Means clustering called affinity clustering for dividing geographic area of Vehicular Ad-hoc Networks (VANETS) into clusters on the basis of infrastructure, type of traffic, and their speeds.

Another prominent technique, known as Hierarchical Agglomerative Clustering (HAC) [13], initially takes every data item as a separate cluster. It then merges the two nearest or similar data items into a single cluster. This step is repeated until either K number of clusters is achieved or until no two clusters have distance smaller than threshold value. It has a number of implementations in the form of single, complete, group-average, and centroid linkages. Lung et al. [16] used HAC in WSN to obtain communication efficiency.

Unlike traditional clustering techniques like K-Means and HAC, spectral clustering is simple to implement through linear algebra [17]. It uses Laplacian matrix derived from the adjacency and degree matrices of the network graph under consideration. Furthermore, with the help of Eigen values and Fiedler vector, the graph is partitioned into K parts. Depending upon graph Laplacian, it has three different versions, that is, unnormalized, normalized symmetric, and normalized random spectral partitioning methods. Along with other clustering techniques, spectral clustering was also used in WSN to cluster nodes in groups [18].

Though all of these clustering techniques are well judged for their efficiency and applications in different areas, however, they have certain limitations; all of them require specification of the number of clusters beforehand while one may not make an intelligent guess in certain cases. Moreover, each of the above techniques can work only when contact information about every node in the network is available. Once this control information, supposedly through flooding, is available, clustering can be performed.

In PSN, after every node knows about its cluster ID and cluster members, it has only to keep track record of its cluster members for intracluster routing. However, in case data is destined to a separate cluster, then source node should know the set of co-cluster members that may deliver
Let $C_s$ be source cluster, $C_d$ be destination cluster, $An$ is adjacency matrix of network nodes, $Ac$ is the adjacency matrix of network clusters.

BridgeNodes($C_s$, $C_d$, $An$)

1. For each node $i \in C_s$ and $j \in C_d$
   1. If $An[i][j]$ is equal to 1 then
      1. PRINT “Bridge nodes from $C_s$ to $C_d$: $i \rightarrow j$”
      2. $Ac[C_i][C_d] := 1$

2. If there is no bridge nodes between $C_s$ and $C_d$
   $Ac[C_s][C_d] := 0$

Algorithm 1: Finding bridge nodes.

Let sender cluster is $C_s$ and destination clusters $C_d$. $Ac$ is adjacency matrix of network clusters

1. Use Warshall algorithm to derive path matrix $P$ for adjacency matrix $Ac$. Where cell $P(i, j)$ contains:
   0, when $C_i$ and $C_j$ cannot access each other directly or indirectly
   $C_z$(ID), when $C_i$ can access $C_j$ via neighbour cluster $C_z$
   $C_j$(ID), when $C_i$ can directly access $C_j$ through its bridge node(s).

2. For each disjoint cluster pairs $C_s$ and $C_d$
   If $P(s, d)$ is equal to 0
      PRINT “No path from $C_s$ to $C_d”$
   Else while $P(s, d) != ID of d$

Algorithm 2: Disjoint_Clusters_Path.

the message early and with surety to the destination cluster avoiding burden on irrelevant links, buffer spaces, and message expiry.

In the next section, we describe a technique for identifying the bridging nodes or paths between every pair of clusters.

3. Bridge Nodes and Paths between Clusters

The message transfer in PSN maintains and exploits the history of encounters among the network nodes for message forwarding decisions [19]. Clustering is a tool to achieve increased benefit from a flat history-based routing algorithm in PSN. It divides the network into small manageable parts so that a node is free from maintaining information about the nodes residing in other parts of the network. Clustering not only reduces burden on network buffers but also saves links from exchanging control information between every pair of network nodes.

The existing clustering techniques, however, fulfill partial requirement for efficient routing, that is, when source and destination belong to the same cluster. What if destination is a member of a separate cluster? In such a situation, every source node should know the community member who ensures delivery of message or the neighboring community which can work as a relay. Algorithms 1 and 2 state the process of selecting bridge nodes between directly accessible clusters and path between indirectly reachable clusters.

In Algorithm 1, $An$ is the adjacency matrix showing contact relationship between every pair of nodes in the network. In case the frequency of encounters and contact durations between a pair of nodes is higher than the threshold value, their mutual value in the adjacency matrix is equal to one; else it is zero. In case the two clusters cannot access each other directly, then path via those neighboring clusters should be identified which can work as relay for forwarding message straight towards the destination cluster. Algorithm 2 extracts path between disjoint clusters.

Algorithms 1 and 2 along with a clustering algorithm are used for routing messages both inside a cluster and between two clusters. Next is the issue of how to choose an appropriate clustering technique for PSN. The next section discusses the parameters on the basis of which accuracy of the clustering techniques is evaluated.

4. Quality Functions and Datasets

An efficient clustering technique depends upon the time required for deriving clusters from data items, the cohesiveness among cluster members, and distance between two different clusters. The following parameters are used for deciding the best clustering technique in the clustering algorithms for a social network:

Complexity. It is the time taken by the clustering algorithm to extract clusters from a dataset.

Modularity. It is based on the idea that a null model is not expected to have a cluster structure, so the possible existence of clusters is revealed by the comparison between the actual density of edges in a subgraph and the density one would expect to have in the subgraph of the null model. Large positive values of modularity indicate good partitions. It is always less than 1 and
can be negative as well, in which case it means that graph has no community structure.

Expansion. The average number of edges per node going away from its cluster is termed as expansion of the cluster. A lower expansion value means that cluster is more self-centered and well structured.

Conductance. It is the ratio of the number of edges pointing outside a cluster to the total number of edges inside a cluster plus those going outside that cluster. The lower the value of a conductance quality function is, the better the clusters are considered.

In this paper, we have used four different datasets for testing various clustering techniques. All of these datasets present relationship levels among people of a certain group:

(1) Zachary’s Karate Club. This data set relates to a popular social network of a university karate club given by Zachary in 1977 [20]. The members of the club are considered as nodes and ties between any two as edges. The dataset consists of 34 nodes and 78 undirected edges.

(2) St. Andrew Town. In this data set, 27 T-mote invent devices were distributed among 22 undergraduate students, 3 postgraduate students, and 2 staff members of University of St. Andrews (available at http://crawdad.org/st_andrews/sassy/20110603/). Participants were asked to carry the devices for 79 days. T-motes came up with a collection of encounters throughout St. Andrew’s town. We, however, have taken into consideration encounters among the participants only.

(3) Les Misérables. This dataset represents the coappearance of a novel’s characters in similar scenes [21]. It consists of around 72 characters.

(4) Third-Grade Students. In 1993, Parker and Asher collected children’s friendships record of third grade in an elementary school. Each child was asked to choose his/her very best friend, three best friends, and any number of friends in his/her class [22]. This class consisted of 22 children.

### 5. Results and Discussion

All the algorithms were run on each of the four datasets for two, three, and four numbers of clusters, respectively. The simulations are performed in a Java based simulator developed for this purpose. Snapshots are taken from the simulation after running Girvan and Newman clustering algorithm [11] to extract four clusters from Karate Club dataset. Table 1 shows the clusters and their members that are extracted by the clustering algorithm. Similarly, Table 2 shows the bridges between clusters derived from the Karate Club using Girvan and Newman algorithm. In Table 3, we can see that, in terms of time complexity, HAC is order $n$ faster than the rest of the clustering algorithms. During the simulation it was found that all three spectral algorithms provide equivalent results. Similarly all the four types of HAC algorithms behave mostly alike with group-average scheme performing sometimes best, in metrics other than complexity, and single linkage sometimes worst in the HAC schemes. Apart from complexity KL performs well though its efficiency depends upon the initial selection of group members.

The average results of modularity, expansion, and conductance from all the simulation runs are normalized to a single value, that is, 4. The one that shows the best result in a quality function is given the highest value. The average result can be easily analyzed from Figure 1.

Girvan and Newman’s algorithm outperforms other algorithms in terms of both expansion and conductance; however, it does not work well in terms of modularity. KL and spectral algorithms behave similarly with KL achieving the highest value in modularity. Other than time complexity, HAC schemes cluster nodes poorly in social networks.

Since modularity is most commonly considered as a decision-maker for terming clusters as good or bad, therefore, according to the above output KL, followed by spectral and then Girvan and Newman schemes will be considered suitable for dividing a social network into groups.

### 6. Conclusion

Pocket Switched Networks (PSN) comprise of portable wireless devices owned by human beings. The network uses embedded low range wireless technologies for message communication. In order to enable efficient routing by decreasing undue stress on network resources, the network needs to be divided into clusters. Every node is limited to maintain contact record of co-community members only which enhances intercluster routing performance.
Table 3: Value of quality functions after running each of the discussed routing algorithms on Karate Club dataset for deriving four clusters.

| Algorithm                     | Expansion | Conductance | Modularity                          | Complexity                  |
|-------------------------------|-----------|-------------|-------------------------------------|-----------------------------|
| HAC/single linkage            | 7.1       | 1.0         | 0.07683673469387563                 | \( O(n^{2}) \)             |
| HAC/complete linkage         | 5.210851648351648 | 1.0 | 0.19734693877550966                 | \( O(n^{2}) \)             |
| HAC/group-average linkage     | 6.790178571428571 | 1.0 | 0.17826530612244817                 | \( O(n^{2}) \)             |
| HAC/centroid linkage         | 6.490625  | 1.0         | 0.1425510204081627                 | \( O(n^{2}) \)             |
| Normalized Spectral Clustering | 2.133531746031746 | 0.3984638047138047 | 0.36561224489795785 | \( O(n^{2}) \)             |
| Unnormalized Spectral Clustering | 2.133531746031746 | 0.3984638047138047 | 0.36561224489795785 | \( O(n^{2}) \)             |
| Normalized Symmetric Spectra Clustering | 2.133531746031746 | 0.3984638047138047 | 0.36561224489795785 | \( O(n^{2}) \)             |
| KL Algorithm                 | 1.1791666666666667 | 0.2872055811571941 | 0.5078571428571413 | \( O(\text{number of iterations} \times n^{2}) \) |
| Girvan and Newman            | 1.4260416666666667 | 0.2872055811571941 | 0.5078571428571413 | \( O(n^{2}) \)             |

Figure 1: Normalized average result.

The routing efficiency highly depends upon selecting the right clustering algorithm. After evaluating different state-of-the-art clustering techniques, we found that KL followed by spectral algorithms derives high quality clusters from social networks.

To enable intracluster routing, the selected clustering schemes should be augmented with discovering paths between cluster pairs. This paper presented an algorithm for finding bridge nodes between directly accessible clusters. Similarly, another algorithm is given to find cluster paths between indirectly reachable clusters.

The above schemes can cluster a set of data points only when the global knowledge about network graph is available. In a social network, a single centralized node must maintain the relationship information between all pairs of network nodes to extract subgroups accurately. In case no node has complete connectivity information about the whole network, the discussed schemes may not work. This situation mostly happens in disrupted networks where nodes are only aware of their familiar nodes. Moreover, changes occur in relationships in the long run which also need to be dealt with. Our future aim is to discover a dynamic and distributed clustering algorithm that may be applicable when a network faces difficulty in collecting updated global connectivity information.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] A.-K. Pietilainen and C. Diot, “Social pocket switched networks,” in Proceedings of the IEEE INFOCOM Students Workshop (INFOCOM ’09), pp. 1–2, Rio de Janeiro, Brazil, April 2009.
[2] D. Hua, X. Du, Y. Qian, and S. Yan, “A DTN routing protocol based on hierarchy forwarding and cluster control,” in Proceedings of the International Conference on Computational Intelligence and Security (CIS ’09), pp. 397–401, IEEE, Beijing, China, December 2009.
[3] Y. Tao and X.-F. Wang, “Adaptive clustering hierarchy routing for delay tolerant network,” Journal of Central South University of Technology, vol. 19, no. 6, pp. 1577–1582, 2012.
[4] H. Dang and H. Wu, “Clustering and cluster-based routing protocol for delay-tolerant mobile networks,” IEEE Transactions on Wireless Communications, vol. 9, no. 6, pp. 1874–1881, 2010.
[5] T. Weninger and W. H. Hsu, “Web content extraction through histogram clustering,” in Proceedings of the International Conference on Artificial Neural Networks in Engineering (ANNIE ’08), vol. 18, St. Louis, Mo, USA, November 2008.
[6] W. Singh, Robustness of three hierarchical agglomerative clustering techniques for ecological data [M.S. thesis], University of Iceland, Reykjavik, Iceland, 2008.
[7] M. Basavaraju and R. Prabhakar, “A novel method of spam mail detection using text based clustering approach,” International Journal of Computer Applications, vol. 5, no. 4, pp. 15–25, 2010.
[8] Y. Liu, Q. Gu, J. P. Hou, J. Han, and J. Ma, “A network-assisted co-clustering algorithm to discover cancer subtypes based on gene expression,” BMC Bioinformatics, vol. 15, no. 1, article 37, 2014.
[9] F. Dresslor, “Clustering,” in Self-Organization in Sensor and Actor Networks, F. Dresslor, Ed., John Wiley & Sons, Chichester, UK, 2007.
[10] P. Hui, E. Yoneki, S. Y. Chan, and J. Crowcroft, “Distributed community detection in delay tolerant networks,” in Proceedings of the 2nd ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch ’07), ACM, Kyoto, Japan, August 2007.

[11] 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, vol. 99, no. 12, pp. 7821–7826, 2002.

[12] P. Costa, C. Mascolo, M. Musolesi, and G. P. Picco, “Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks,” IEEE Journal on Selected Areas in Communications, vol. 26, no. 5, pp. 748–760, 2008.

[13] M. H. Dunham and S. Sridhar, Data Mining: Introductory and Advanced Topics, Pearson Education, New Delhi, India, 1st edition, 2006.

[14] S. Jerusha, K. Kulothungan, and A. Kannan, “Location aware cluster-based routing in wireless sensor networks,” International Journal of Computer and Communication Technology, vol. 3, no. 5, pp. 1–6, 2012.

[15] M. Bhaumik, S. DasGupta, and S. Saha, “Affinity based clustering routing protocol for vehicular ad hoc networks,” Procedia Engineering, vol. 38, pp. 673–679, 2012, Proceedings of the International Conference on Modelling, Optimization and Computing, April 2012, Tamil Nadu, India.

[16] C. H. Lung, C. Zhou, and Y. Yang, “Applying hierarchical agglomerative clustering to wireless sensor network,” in Proceedings of the International Workshop on Theoretical and Algorithmic Aspects of Sensor and Ad-Hoc Networks (WTASA ’07), pp. 97–105, Miami, Fla, USA, June 2007.

[17] U. von Luxburg, “A tutorial on spectral clustering,” Statistics and Computing, vol. 17, no. 4, pp. 395–416, 2007.

[18] A. Jorio, S. E. Fkihi, B. Elbhiri, and D. Aboutajdine, “A new clustering algorithm in WSN based on spectral clustering and residual energy,” in Proceedings of the 17th International Conference on Sensor Technologies and Applications (SENSORCOMM ’13), Barcelona, Spain, August 2013.

[19] K. Massri, A. Vernata, and A. Vitalletti, “Routing protocols for delay tolerant networks: a quantitative evaluation,” in Proceedings of the 15th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, ACM, Paphos, Cyprus, October 2012.

[20] W. W. Zachary, “An information flow model for conflict and fission in small groups,” Journal of Anthropological Research, vol. 33, pp. 452–473, 1977.

[21] D. E. Knuth, The Stanford GraphBase: A Platform for Combinatorial Computing, Addison-Wesley Professional, 1st edition, 1993.

[22] J. G. Parker and S. R. Asher, “Friendship and friendship quality in middle childhood: links with peer group acceptance and feelings of loneliness and social dissatisfaction,” Developmental Psychology, vol. 29, no. 4, pp. 611–621, 1993.