International Journal of Computer Applications (0975 – 8887)
Volume 103 – No.4, October 2014

A Clustering Algorithm in Complex Social Networks

Veera Nagaiah Maddikayala
Cambridge Institute of Technology
Bangalore, Karnataka, India

R Chandrasekhar
Atria Institute of Technology
Bangalore, Karnataka, India

ABSTRACT
Complex networks are real graphs (networks) with non-trivial topological features. The empirical study of real-world networks like computer networks and social networks gives insights into the structures and properties of such networks. Identification of community structure is one of the important problems in social networks. Tightly knit group of nodes (Cluster) characterized by a relatively high density of ties (links) tend to be greater than the nodes that have average probability of ties randomly established [8][16]. In this paper a novel clustering algorithm is developed in complex social networks to detect the communities with close relations where in, everybody is aware of every other in their group called cluster. Determining such groups is the main concern of this paper. Some of the social networks are online Facebook, LinkedIn, Twitter and day today socializing. Graph Theoretic approach is followed for finding the clusters. Perfect graph structures are investigated in the complex social networks.

General Terms
Networks, Clusters, Graphs

Keywords
Complex social networks, scale-free networks, perfect graphs, social clusters, independent set, and cliques.

1. INTRODUCTION
Complex networks are not formally defined but are characterized by dynamically changing big networks which are backbones of complex systems. The origin of complex networks can be looked back with the remarkable work on random graphs by Erdős and Rènyi [6][7]. The inspiration to the domain of complex networks is from the real world networks like social networks, information networks, Technological networks and biological networks. The important properties seen in the complex networks [4] are small world effect [15], transitivity, degree distribution, network resilience, degree correlations, community structure and mixing pattern and network navigation.

Social networks are modeled in various ways. Among these models, the random graph model of Paul Erdős and Rènyi (ER), the Small-World Model of Watts and Strogatz (WS), and Scale-free networks of Barabàsi and Albert (BA) [3] are nearer to the real world phenomenon.

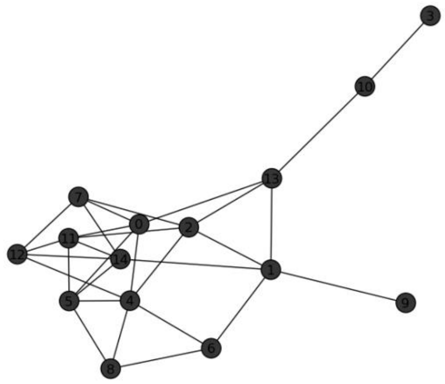

Fig. 1 Erdős and Rènyi graph with nodes n = 15, probability p = 0.3.

ER graph (network) is defined by the number of vertices *n* and the probability *p* that an edge between two given vertices exists (Figure.2). The **expected degree** of a vertex in the network is

$$\langle k \rangle = p\,(n - 1)$$

Watts and Strogatz [14] have shown that the degree distribution for small-world networks is similar to that of random networks (Figure.2) with a peak at

$$\langle k \rangle = 2\,l$$

where l is the neighborhood. According to of Barabàsi and Albert, instead of the vertices of these networks having a random pattern of connections, some vertices are highly connected while others have few connections exhibiting scale free behaviour [9]. The degree distribution P follows a power law for large k, $P\ (k) = k^{-\gamma}$ where γ is an integer which depends on the type of the network.

2. CLUSTERS IN COMPLEX SOCIAL NETWORKS
Let G (V, E) be a complex network with V nodes and E edges. In this paper G in particular is an online Social Network like Facebook, LinkedIn, Twitter or any Social network of daily face to face interactions. Being complex, the social network has very large number of nodes in terms of thousands or even in millions.

This network can be divided into the induced subnets (sub-networks) with nodes having common interest called social clusters. .

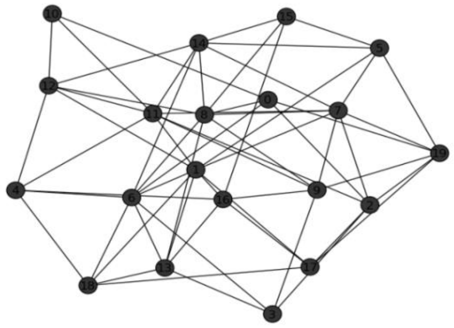

Fig 2. Watts and Strogatz graph with nodes *n* = 20, probability *p* = 0.7.

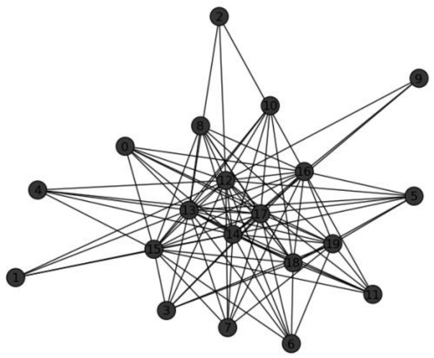

Fig 3. Barabàsi and Albert graph with nodes n = 20, number of edges to attach from a new node to existing nodes m = 12.

A social cluster C is a subnet of G i.e $C \subseteq G$ with at least one common property. Now a Social network G can be represented as $G = C_1 \cup C_2 \cup C_3 \cup \ldots \cup C_r$, where r is the number of clusters in G. One of the typical problems is to find various clusters in a given complex social network. Graph theoretically social clusters can be represented as complete sub-graphs (cliques) of a complex random graph.

A local graph clustering algorithm finds a solution to the clustering problem without looking at the whole graph [17]. Such algorithms are useful for handling massive graphs, like social networks and web-graphs [13] in linear time. In this paper a clustering algorithm with perfect graph structure of a given probability is considered. Once the clusters in the social networks are identified through the proposed algorithm they can be interpreted and analyzed in studying the social behaviour online and offline.

Terminology:

G: A complex social network.
H: An induced network of G

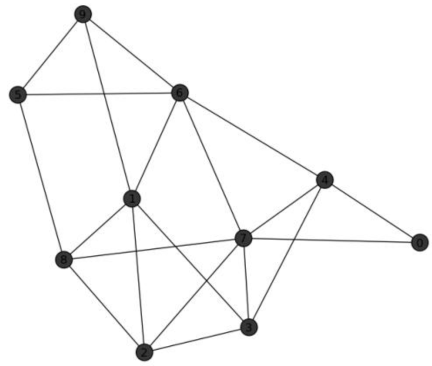

Fig 4. Watts Strogatz small world network n = 10, k = 4, p = 0.7

χ(G): The minimum independent set needed to cover the nodes of G.

Θ(G): The cluster cover number of G, the minimum number of complete sub-networks of G needed to cover the nodes of G

ω(G): The cluster number, the maximum number of mutually adjacent nodes, that is the size of the largest complete cluster of G.

α(G): The size of the largest independent (stable) set of nodes.

S: Stable set which meets all the maximal cliques in G.

N(v): Set of neighbours of v.

Definition: A complex social network G is said to be perfect network if and only if the following conditions (i) and (ii) hold

good [12].

(i) $\alpha(G) = \Theta(G)$ and

(ii) $\omega(G) = \chi(G)$

Examples of perfect social networks are Complete networks, Bipartite networks, Triangulated networks, Meyniel Network. Meyniel Network is a network with every odd cycle of atleast length five have atleast two chords [11]. A network is called strongly perfect if each of its induced subnet H contains a stable set of nodes which meets all the maximal clusters in H.

3. THE PROPOSED CLUSTERING ALGORITHM

Given a random social network, the clusters as cliques (maximally complete sub-graphs or sub-networks) can be found [5]. Clusters of a given network graph G can be identified as an induced set of nodes (vertices) of G. The Watts Strogatz small world network of size n = 10, k = 4, p = 0.7 as shown in Figure 4 is a good example for the social network. G Ravindra [11] obtained an efficient algorithm to find either a starter in some induced sub-network of a network or an independent set of nodes which meets all the maximal cliques in the network. If independent set is found then the network is a strongly perfect and perfect network [2] as well.

Here starter in a network means, in any complex network, a cycle w v_0 v_1 . . . v_k such that:

(i) v_0 is adjacent to none of the vertices v_2, v_3, ... v_k

(ii) w is not adjacent to v_1

(iii) there exists some stable set of nodes S, containing v_1 and v_k that meets all the maximal clusters in G - v_0.

Definition: A Complex network G is called strongly perfect if each of its induced sub-networks H contains an independent set which meets all the cliques (clusters) in H.

Given a complex network, a $K_{1,3}$ free network is always a strongly perfect network and there is always a maximum independent set of nodes.

The maximum independent set of nodes for the Barabàsi and Albert graph [1] with nodes n = 30, edges m = 10 (Figure 5) and The Watts Strogatz complex social network (Figure 4) are found.

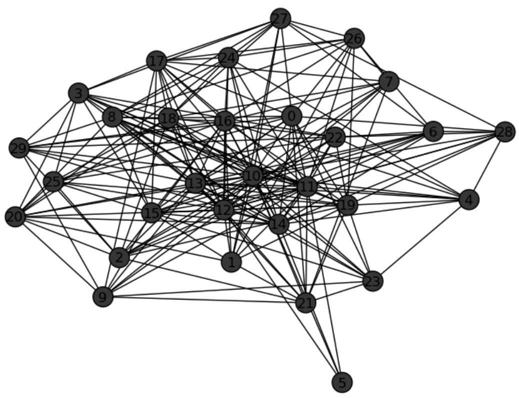

Fig 5. Barabàsi and Albert graph with nodes *n* = 30, number of edges to attach from a new node to existing nodes *m* = 10.

Algorithm to find clusters in a given complex social network

Input: Complex social network G
Output: The cluster heads of the complex social network S (Independent set)
Step1: START
Step2: Read the complex social network G
Step3: Find starter
Step4: If (starter found)
  print "The Complex Network is not a Strongly Perfect one"
Step5: else
  Find an induced sub-network H of G
Step6: Choose a node t in G and find H = G - t - N(t)
Step7: If H = Φ ; then S = { t }
  print S
  else
   choose v_0 in H
   S = { v_0 }
 for each x in G - v_0
 Begin
   F = G - v_0 – N(v_0) ∪ { t }
   H = H - v_0 - N(v_0)
   S = S ∪ { x }
  if nodes v_1, v_2 ε G - v_0 are not found such that
   v_1 ε N(v_0) ∩ S and v_2 ε N(v_1) ∩ F
  then
  output " S - N(v_0) meets all maximal clusters in G"
  else if nodes w, z ε G – { v_0 v_1 v_2 } are not found
   such that w ε N(v_0) - N(v_1) and z ε N(w) ∩ F ∩ S
  output "The stable set S meets all maximal clusters in G"
 else
   continue
 End
Step8: Print the stable set S which meets all the clusters in the Complex Network.
Step9: STOP

4. SIMULATION RESULTS

The algorithm is run on various types of social networks to find the maximum independent set and the results are listed in the Table-1. The first column shows the type of social network, second shows number of nodes and the third shows the maximum independent set of respective type of social network. the numbers shown in the independent set column represent the nodes in the network that form the stable set. After finding such a set, strongly knitted groups of communities are identified in the Network. Given a node (member) in the complex network, it can be found in one of the clusters of that network. Usually the maximum independent set and related problems are formulated as nonlinear programs. It is observed that as the growth of the cardinality of the independent set is linear with the number of growing nodes in the network (Figure 6).

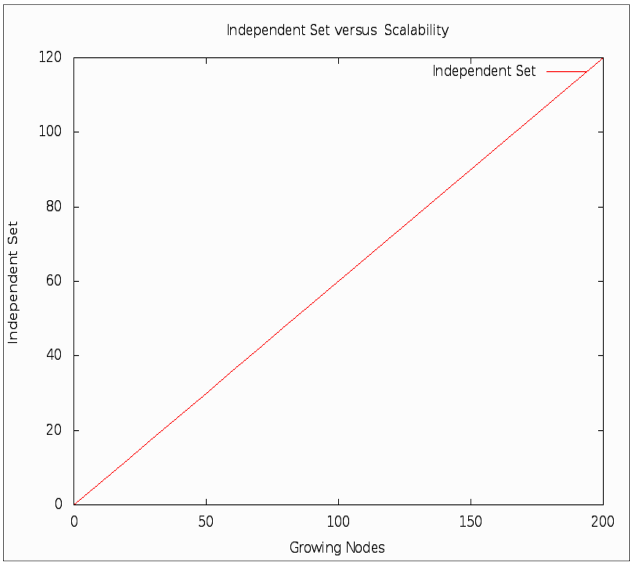

Fig 6. Relation between the independent set and the growing nodes in complex network

Simulation results of the algorithm are tabulated in Table-1, Table-2 and Table-3. The plot of the independent sets(number of clusters) verses the growing nodes is shown in Figure 6 and the relations between the independent sets and the growing links are shown in Figure 7.

Table 1. Maximum Independent Sets for different Social Networks

Type of Social Network	Number of Nodes	Maximum Independent Sets
Erdős and RènyiNetwork	50	{ 12, 27, 46, 37 }
Barabàsi Albert Network	70	{ 42, 63, 35, 29 }
Newman Watts Strogatz Network	60	{ 51, 39, 33, 3, 6, 36, 54, 15, 24, 43, 48, 59, 29, 12, 19 }
Random Regular Network	40	{ 2, 34, 11, 31, 0, 19, 33, 5, 20, 10, 35, 6 }
Karate Club Network	30	{ 18, 4, 30, 13, 27, 16, 29, 20, 22, 12, 21, 15, 25, 19, 17, 9, 14, 7, 28, 11 }

Table 2. The cluster number for two types of networks

Growing Links	Scale Free Networks	Random Networks
50	38	457
90	26	430
120	35	416
150	77	408
180	119	400
220	220	393
250	250	353
300	300	329
350	350	322
400	400	306
499	499	282

Table 3. Effect of Scaling on Clusters

Number of Nodes	Number of Clusters
50	26
100	49
150	78
200	105
250	132
300	151
500	275
1000	546
5000	2741

It can be observed that the identification of the clusters depends on the type of the social network. In scale free networks the clusters grow as the links grow, where as the number reduces in random networks.

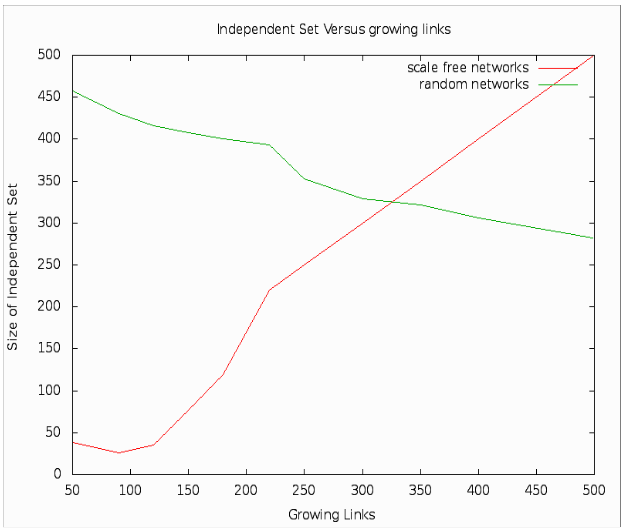

Fig. 7. Clusters with growing links in the Complex Social Network

5. APPLICATIONS OF THE ALGORITHM

A complex social network can be treated as a perfect network, if there exists a stable set of nodes which meets all the maximal clusters in the network. There are many areas [10] in which this algorithm can be applied. Some of the examples are in finding the peer groups on the Facebook, LinkedIn, Twitter and Blogs in the internet of common interest like research groups, developers on a particular domain, web pages of on specific information etc. The leaders in the society can be identified with their clusters. Departments in an organization, food webs make clusters in their respective networks. The algorithm can also be applied in many more areas like information retrieval, classification theory, economics, scheduling, experimental design and computer vision.

6. CONCLUSION

An algorithm to find the clusters in Complex Social Networks like Facebook, LinkedIn, Twitter etc., who are connected through web links is developed. A graph theoretic approach is followed using perfect graphs to build the algorithm. The information in clusters can be used to interpret the dynamics of the group. A cluster of computers in data communication networks participate in processing the data packets and messages. Computer clusters are deployed to improve cost-effectiveness, performance and availability compared to sparsely connected computers. They have a wide range of applicability and deployment, ranging from small business clusters with a handful of nodes to the computer clouds.

7. REFERENCES

[1] Albert-László Barabási and Rèka Albert. Emergence of scaling in random networks. science, 286(5439):509–512, 1999.

[2] C Berge and P Duchet. Strongly perfect graphs. Annals of Discrete Mathematics, 21:57–61, 1984.

[3] L da F Costa, Francisco A Rodrigues, Gonzalo Travieso, and P R Villas Boas. Characterization of complex networks: A survey of measurements. Advances in Physics, 56(1):167–242, 2007.

[4] Bhaskar DasGupta and Devendra Desai. On the complexity of newmans community finding approach for biological and social networks. Journal of Computer and System Sciences, 79(1):50–67, 2013.

[5] E Knuth Donald. The art of computer programming. Sorting and searching, 3:426–458, 1999.

[6] Paul Erdős and Alfréd Rényi. On random graphs. Publicationes Mathematicae Debrecen, 6:290–297, 1959.

[7] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. Bull. Inst. Internat. Statist, 38(4):343–347, 1961.

[8] Paul W Holland and Samuel Leinhardt. Transitivity in structural models of small groups. Comparative Group Studies, 1971.

[9] Ajaz Hussain, Khalid Latif, Aimal Tariq Rextin, Amir Hayat, and Masoon Alam. Scalable visualization of semantic nets using power-law graphs. Applied Mathematics & Information Sciences, 8(1), 2014.

[10] Thilo Kielmann, Henri E Bal, and Sergei Gorlatch. Bandwidth-efficient collective communication for clustered wide area systems. In Parallel and Distributed Processing Symposium, 2000. IPDPS 2000. Proceedings. 14th International, pages 492–499. IEEE, 2000.

[11] G Ravindra. Meyniels graphs are strongly perfect, in Topics on perfect graphs (eds. c berge and v. chvtal). Ann. Disc. Math, 21:145–148, 1984.

[12] G Ravindra and K R Parthasarathy. Perfect product graphs.Discrete Mathematics, 20:177–186, 1977.

[13] Daniel A Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. SIAM Journal on Computing, 42(1):1–26, 2013.

[14] Steven H Strogatz. Exploring complex networks. Nature, 410(6825):268–276, 2001.

[15] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. Sociometry, pages 425–443, 1969.

[16] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world' networks. nature, 393(6684):440–442, 1998.

[17] Yourim Yoon and Yong-Hyuk Kim. Vertex ordering, clustering, and their application to graph partitioning. Applied Mathematics & Information Sciences, 8(1), 2014.