Taxonomy of Community Models and Applications in Complex Networks

1st Abdelhani Diboune
MOVEP Laboratory
U.S.T.H.B
Algiers, Algeria
adiboune@usthb.dz

2nd Hassina Nacer
MOVEP Laboratory
U.S.T.H.B
Algiers, Algeria
sino_nacer@yahoo.fr

3rd Beghdad Bey Kadda
DCS Laboratory
E.M.P
Algiers, Algeria
k.beghdadbey@gmail.com

4th Hachem Slimani
LIMED laboratory
Université de Bejaia
Béjaïa, Algeria
haslimani@gmail.com

***Abstract*—The analysis of complex networks is an interdisciplinary area that combines many fields such as data analysis, graph theory, computer science, etc. It gives an accurate description of many phenomena that imply interactions between different components of complex systems. One of the most important issues in studying such networks is probing community structures. It consists of dividing the network into modules with dense intra-connections, and sparse interconnections. The nodes within such communities share common features and collaborate to execute same task in the network. This paper aims to survey models of underlying communities in complex networks that have been studied by researchers. In addition to that, it gives an insight to the challenging issue of interpretation of underlying communities, according to the fields of their applications and network structures.**

***Index Terms*—Complex network, community detection, collaborative filtering, Sybil defence, web communities, service composition.**

I. Introduction

Complex networks are irregular and dynamic structures that present three main characteristics [1]: the small world propriety, the scale-free connectivity distribution and high clustering coefficient (transitivity). 1) The small world propriety means that the average path length between any two nodes in the network increases languidly (logarithmically) with the network size, 2) the connectivity distribution in complex networks is characterised by a power law 3), finally, complex network has a greater clustering coefficient. According to their application fields, E.M.J. Newman has divided complex networks into four categories [2]: social networks, technological networks, biological networks and information networks. This paper studies information networks which cover three main classes: 1) two-party communication networks (e.g. email and IM), 2) multi-party communication (e.g., bulletin boards and mailing lists), 3) content-sharing sites (e.g., YouTube and Flickr).

By the development of big data and advanced data analytics platforms, huge amount of topological information from archival records (mail servers' logs, online journals, etc.) allowed a precise description of topology and dynamics of very large information networks. However, the amount of data is so large that it is difficult to identify the exact topological structures. In addition to that, the lack of a mesoscopic description makes it hard to depict different autonomous modules in the network and the way nodes collaborate within them. Thus, probing community structures is one of the most important issues in studying these complex networks. Indeed, automatic discovery of network communities appears to be a good solution for revealing the coarse-grained structure of networks which are too complex for users to make sense of at the individual vertices level and understanding the features of nodes just from network topology. Even though, many publications have studied clustering algorithms [3], the network communities have been scarcely investigated from the application point of view. Indeed, till now, there is not a formal consensus for defining either what are communities, or what are the purposes of such coalitions. The common definition of a community given in most publications is a set of vertices which share common features or collaborate to execute same tasks. Another commonly used one is a subgraph in which the number of internal edges is larger than the number of external edges [3].

In this paper, we try to survey taxonomy of community models in complex information networks, their advantages and drawbacks. In comparison to the existing works, the main contribution of this work consists of introducing the problem of interpreting the underlying communities in information networks, according to their application domain.

The reminder of this paper is organised as follows: the next section outlines the existing community models in the literature. Section III classifies network communities according to their applications. Section IV compares existing proposals that use community structures. Finally, section V concludes the paper.

II. Community as a mathematical model

In this section, we explore the different mathematical models used in literature to describe communities in complex networks. These models may be divided into four classes 1) graph model 2) stochastic model 3) data clustering model 4) game theoretic model.

The interconnection network is represented as a weighted graph $G(V(G), E(G), W)$ or unweighed graph $G(V(G), E(G))$. Where $G(V(G)$ is vertices set, $E \in V^2$ is edges set and $W : E(G) \longrightarrow \mathbb{R}^+$ is weight function that associates to each edge (i, j) a real number $w_{i,j}$. In the following, $N(k)$ is the set of neighbours node $k \in V$ and $d_k \in V$ the degree of node k, where

$$d_k = \sum_{k \in N_u} w_{k,i} \tag{1}$$

A. Graph model

A community in network is defined as highly cohesive subgroup of nodes. The cohesiveness can be quantified by two dimensions [4] : completeness and centripetal-centrifugal.

1) Completeness dimension: means that high fraction of pairs in the community are directly linked by the appropriate relation. One of the most popular graph concepts used to define communities are:

a) Clique: maximum complete subgraph of G [5], [6].

b) N-clique: maximal subgraph G' of G such that [6]:

$$\forall u, v \in G' : d_G(u, v) \leq n. \tag{2}$$

Where $d_G(u, v)$ is the distance between u and v in graph G

c) N-clans: n-clan G' of G is an n-clique of G such that

$$\forall u, v \in G' : d_{G'}(u, v) \leq n. \tag{3}$$

Where $d_{G'}(u, v)$ is the distance in graph G'. This model allows a better cohesion [7] within the community and avoids disconnections between nodes.

d) N-club: maximal subgraph of G of diameter n [5].

e) K-plex: subset of k nodes $(k < n)$ in the network such that each node is a neighbour of at least $(n - k)$ other nodes in the subset [7].

2) Centripetal-centrifugal dimension: this class measures how much nodes of the community are related to external nodes. For this, measures of centripetal cohesiveness of a community C have been introduced such as cut-ratio δ_C^{ext}, which is defined as the ratio between the number of external edges K_C^{ext} and the total possible external edges $n_c(n - n_c)$,

$$n = \mid V(G) \mid \frac{K_C^{ext}}{n_C(n - n_C)}. \tag{4}$$

Then, a set of nodes is assumed to be a community if this metric is less than a given threshold ξ. $(\delta_C^{ext} < \xi)$ [4].

B. Stochastic model

In stochastic model, community is considered as subnetwork in which nodes have higher probability to form edges with other nodes within the same community than with the other vertices.

Let $C(V(C), E(C))$ be a subnetwork of $G(V(G), E(G))$. $P_{i,j}$: probability to form an edge between nodes i and j. In this scope, there is three models [4]:

1) Strong stochastic model: $C(V(C), E(C))$ is a strong stochastic community if:

$$\forall i \in V(C) : \forall j \in V(C), \forall k \in V(G) - V(C) : P_{i,j} < P_{i,k}. \tag{5}$$

2) Weak stochastic model: $C(V(C), E(C))$ is a weak stochastic community if:

$$\forall i \in V(C) : \frac{\sum_{j \in V(C)} p_{i,j}}{|V(C)|} > \frac{\sum_{j \in V(G)-V(C)} p_{i,j}}{|V(C)|}. \tag{6}$$

3) Markov chain and random walk model: a random walk is associated to a graph $G(V(G), E(G))$, which is a path, such that at each step t an outgoing edge (i, j) from node i to node j is chosen with probability:

$$P_{i,j} = \frac{w_{i,j}}{d_k}. \tag{7}$$

The probability of finding a walker at node $j \in V(G)$ at t is:

$$p_j(t) = \sum_{k \in N(j)} P_{k,j}\ p(t-1). \tag{8}$$

C is a community if

$$\forall j \in V(C), \forall k \notin V(C) : P_j(t) > P_k(t). \tag{9}$$

C. Data cluster model

Let N be a network with a set of nodes $p_1, p_2, \ldots, p_n$ each node p_i is considered as a vector of features

$$p_i = [f_{i1}, f_{i2}, \ldots, f_{im}]. \tag{10}$$

Clustering is defined as a set of modules $C_1, C_2, \ldots, C_k$ that optimise a specific criterion computed in function of $d(p_i, p_j)$ which is a distance function that determines similarity between each pair of data points p_i, p_j. Every cluster C_i is considered as a community or a coalition.

D. Game theoretic model

Many authors have proposed solutions of depicting community structures based on game theory.

Every game $\Gamma = (N, (S_i)_{i \in N}, (U_i)_{i \in N})$ can be described using three parameters:

- set of players (dicsion makers) $N = 1, 2, \ldots, n$ indexed by i which could be agents, nodes, users, etc.
- The strategy space $(S_i)_{i \in N}$, such that a strategy is a set of nodes communities. For each player i, $S_i = \{S_{i1}, S_{i2}, \ldots, S_{im}\}$ is a set of his strategies.
- $U_i : S_i \mapsto R$ is the utility payoff function, which is the expected income player i gets when $S = \{S_1, S_2, \ldots, S_n\}$ is implemented.

The game can be cooperative or non-cooperative.

1) Non-cooperative game theory model for community detection: in non-cooperative game, the solution concept used to derive good communities is Nash equilibrium; a strategy $S = \{S_1, S_2, \ldots, S_N\}$ is said to be Nash equilibrium if $\forall i \in N, S_i$ is the best response to S_{-i}. i.e. At Nash equilibrium, no player tries to change his strategy.

2) Cooperative (coalitional) game theory model for community detection: in cooperative game theory, a community is a local coalition of players in the network to maximise a payoff (e.g. local modularity). A player chooses a partition according to a Pareto dominance strategy over his payoffs. Then, some communities are split and some neighbouring communities are merged until the Nash equilibrium is reached.

III. Application fields of Communities in complex networks

Even though many works use the general and intuitive definition that considers the community as a module in which the number of intra-community edges is much higher than the number of inter-community edges, it is still confusing and does not define the common features shared between nodes and the task insured by the community in the network. In this section, we review the most common community application domains found in literature, we depict four main fields: 1) collaborative filtering 2) Sybil defence 3) service composition 4) web communities.

A. Communities in collaborative filtering

Collaborative filtering, uses communities to recommend items or services which are liked by the users with similar tastes [37]. Indeed, collaborative filtering has two main subclasses, model based and memory-based. The model-based method finds a model that describes the purchasing behaviour of users and, predicts the future ratings of items. The memory based is classified into user-based and items-based collaborative filtering. While user-based collaborative filtering searches for sets of users who have similar tastes as the target user and computes the rating prediction, item-based collaborative filtering aims to recommend an item based on the ratings of similar items liked by the target user. The collaborative filtering uses clustering methods in order to detect underlying subgroups of users, items [15]–[17], [38]–[40] or web services [12], [41], [42] highly correlated. The communities are generally represented in stochastic model [13] or as data clusters [17], [38]–[40].

B. Communities in Sybil defence

Complex networks are vulnerable to deceptive attacks, which consist of cloning legitimate profile details in order to create fake identities and compromise network functions. Indeed, fake identities can alter the trust of other honest users, suppress content, affect rankings, compromise DHT routing and causes Byzantine failure [43]. Recently, there has been an increasing interest in defending against Sybil attacks (deception attack), in a complex network, especially online social network. Research works in this field are classified into two main classes 1) content based methods and 2) community-based methods [44]. The first class, is mainly based in analysing content and behaviour of communities, while the second is based on tracking evolution of communities in the interconnection graph in order to identify anomalies. In cluster-based Sybils defence approaches, communities are mostly modelled either by graph model or by stochastic models. In graph model, the goal is to depict modules from the network which exhibit distinguishable topological structures from honest nodes [18], [19], [21]. In stochastic model, the goal is to find clusters that do not show fast mixing characteristic [20], [22], [23], [45].

C. Service composition communities

In the web, users request for services which may be ensured by a single service or a composition of multiple services [28]. Service composition faces many challenges, one of the most important is the quality of service. Indeed, given a set of services, the issue consists of finding an optimal cluster of services that collaboratively execute the requested task with high quality of service [28]. Another important issue is security, indeed the collaboration of many services for a better QoS should not come at the cost of the security and trustworthiness dimension [46]. In composition environments some service providers may collect users confidential information from his component services and shares it with other service providers who couldnt always be trustworthy. Thus, the composition environment should be able to detect trust circles for each web service. In this context, many works have studied the problem of detecting clusters of web services. These clusters are mainly modelled either as densely connected subgraphs [24], [25], [27], [47] or highly correlated set of data points [26], [28], [29].

D. Web pages communities

In many publications, web communities are considered as a set of web pages with the same topic [48], a set of related pages to a core URL [49]. i.e. pages treating the same topic as the original page or a group of similar web user session [33]–[35], [50]. Since of decentralised and chaotic generation of content, these communities have dynamic behaviours: which induce highly dynamic emerging cyber-communities [49]. To measure similarities between web pages, many works used the hyperlink-structures of the web [31], [48], [51], [52]: the main idea is that pages belonging to the same communities tend to be more frequently co-cited. Another method consists on finding subnetwork with fast mixing characteristic which is one of the main characteristics of the web induced by the small world characteristic. Some other works use the concept-based similarity of web pages contents. Others used both link information and the content for categorising web pages.

IV. Comparative study

In section II we presented the different mathematical models of communities. Even though these models can be used interchangeably, every model has its advantages and drawbacks.

Graph models give a clear insight about the topological structure of communities, however, the major drawback of these models is that detecting communities is computational expensive (finding cliques, n-clubs, n-cliques is NP-complete problem), another drawback is that tracking the evolution of communities is time consuming because it consists of tracking snapshots of the network at every interval τ and depicting evolution of the community by finding similar communities at each instant.

Compared to graph models, detecting communities in stochastic models is less time consuming. Indeed, in a random walk model the stationary distribution is reached after n steps, such that $n \leq$ diameter of the network. According to the small world characteristic of complex networks, the diameter of network increases logarithmically with its size (in practice $6 \leq n \leq 20$).

Data clustering model takes into consideration both links and node features in the network. However, in most cases algorithms used to detect communities in data cluster models assume that the number of clusters is known which is unrealistic in real applications. In this context, even though some methods are proposed to compute the number of clusters (e.g. applying spectra on non-backtracking and flow matrix) the major drawback of these approaches still high computational complexity.

Game theoretic model is efficient in tracking evolution of communities, because instead of depicting them from the scratch at each time step, it updates the clustering of nodes found in the previous step at every time a set of players change their strategies. However, a key limitation of such strategy is slow convergence in large networks.

In Section III, we cited different community definitions according to their application domains which include collaborative filtering, Sybil defence, service composition and web communities. Even though for every application many methods have been proposed using traditional content-based analysis, community detection based methods have many advantages over them.

1) It accelerates computational time, indeed, the only users analysed by the collaborative filtering algorithms are the ones belonging to the same community (cluster) as the targeted user.
2) Collaborative filtering uses social network information to deal with data sparsity, and cold start issues which means lake of ratings induced by adding new items to the network, and new subscription of users.
3) Network clustering can enhance recommender systems by using social network information. Indeed, users belonging to the same communities tend to have similar behaviours and rely on their connections for choosing items. Once communities are created, predictions for a user can easily be made by aggregating the opinions of other users in the same community.

In Sybil defence, community-based method helps to solve both problems of content-based approaches:

1) Most anomaly behavioural-based detection algorithms are based on analysing the text content (Key-words) in interactions between users. However, many attacks use non-dictionary words in their communication like images and videos in spam messages.
2) Content-base methods require high computations and behaviour-based algorithm allows fake accounts to stay in the network for a while this induces the creation of more connections between Sybil and honest accounts and make traceability of fake account difficult.

Community detection on the web can help improve the effectiveness of ranking web pages by the research engine, as the pages belonging to the same communities will probably match the same research request.

Finally, table I presents a comparative study of community characteristics according to the following criteria: community model, type of nodes, edge weight and metrics used to compute the weight.

To probe communities in the network, we need metrics to compute similarities between nodes. Nodes (users, web services, web pages, etc.) are modelled by n-dimensional vectors which could be vectors of features [35], vectors from the adjacency matrix that describes egocentric network of the node, or rating vector that describes preferences of the user [8] [14].

In some cases, concept-based similarity is used to compute the semantic similarity between entities in the web [28]. Firstly, the ontology tree should be computed. Then, given two concepts c_1 and c_2, the distances between each of them and their common parent p in the tree are computed, that is $d(p, c_1) = d1$, $d(p, c_2) = d_2$. After that, distance between the common parent p and the root is computed d_{root}. Thus, the similarity between c_1 and c_2 can be computed by the formula:

$$sim(c_1, c_2) = \frac{2\ d_{root}}{d_1 + d_2 + 2d_{root}} \tag{11}$$

Let U, V be two n-dimensional vectors representing nodes in the network, U_i, V_i are respectively the ith components of U and V. $\bar{U}$, $\bar{V}$ are average vectors such that:

$$\forall i : \bar{U_i} = \frac{1}{n}\sum_{j=0}^{n-1} U_j,\ \bar{V_i} = \frac{1}{n}\sum_{j=0}^{n-1} V_j \tag{12}$$

The different metrics used in literature to compute the similarity between U and V are described in table II.

V. Conclusion

Network analysis is one of the most important fields in both social computing and big data. In many scenarios, community detection serves as the backbone of network analysis. The two past decades have witnessed the prosperity of the research on community detection in complex networks. In this paper, a thorough review of available community models in complex information networks was presented. The main goal here is firstly to categorise these different mathematical models, study their characteristics, then, analyses their applications in real networks. It is expected that network analysis scope will continue its expansion. The methods and theories that work for community detection will be useful for many networks open issues such as recommender systems' optimisation, network resource allocation, anomaly detection, network ranking, network security, sentiment analysis on networks, and so forth.

TABLE I
COMPARATIVE TABLE OF COMMUNITY CHARACTERISTICS IN COMPLEX INFORMATION NETWORK

Application	Work	Date	Nodes	Links weight	Metric	Community model
Collaborative filtering	[8]	2014	User	Similarity	Pearson correlation	Stochastic
	[9]	2015	User	Trust	Cosine similarity	Data clusters
	[10]	2015	User	Similarity	Bhattacharya coefcient	Data clusters
	[11]	2015	User	Similarity	Hamming distance	Data clusters
	[12]	2016	Web service	Similarity	Pearson correlation	Data clusters
	[13]	2016	User	Probability of belonging	Probabilistic measure	Stochastic
	[14]	2016	User	Trust	Jaccard coefficient	Data clusters
	[15]	2017	User	Similarity	Pearson correlation	Data clusters
	[16]	2017	User	Trust	Pearson correlation	Data clusters
	[17]	2017	User	Similarity	Jaccard Similarity	Data clusters
Sybil communities	[18]	2014	Accounts	Unweighted (trust)	-	Graph model
	[19]	2015	IP addresses	Unweighted (trust)	-	Graph model
	[20]	2016	Accounts	Trust	Power iteration method	Stochastic
	[21]	2016	Accounts	Unweighted (trust)	-	Graph model
	[23]	2017	Account	Unweighted (trust)	-	Stochastic
Service composition	[24]	2012	Cloud services	Distance	Euclidean	Graph model
	[25]	2013	Operations	Similarity	Number of interactions	Graph model
	[26]	2014	Web services	Compliance	-	Data cluster
	[27]	2015	Inputs/Outputs	Similarity, operation	-	Graph model
	[28]	2017	Cloud services	Similarity	Concept-based similarity	Data cluster
	[29]	2018	Web services	Structural similarity	SimRank	Data cluster
Web communities	[30]	2014	Web document	Similarity	Cosine similarity	Data cluster
	[31]	2014	Web 2.0 items	Similarity	-	Graph clusters
	[32]	2014	Web pages	Similarity	-	Data clusters
	[33]	2015	User sessions	Similarity	Hamming distance	Data cluster
	[34]	2016	User sessions	Similarity	Euclidean distance	Data clustering
	[35]	2017	User session	Similarity	-	Data clusters
	[36]	2017	Web images	Correlation	Cosine similarity	Stochastic

TABLE II
COMPARATIVE TABLE OF SIMILARITY METRICS.

Metric	Formula	Works
Pearson correlation	$Per(U,V) = \frac{\sum_{i=0}^{n-1}(U_i-\bar{U})(V_i-\bar{V}_i)}{\sqrt{\sum_{i=0}^{n-1}(U_i-\bar{U})^2}\sqrt{\sum_{i=0}^{n-1}(V_i-\bar{V})^2}}$	[8], [12], [15], [16]
Cosine similarity	$Cos(U,V) = \frac{\vec{U}.\vec{V}}{\|\vec{U}\|.\|\vec{V}\|}$	[9], [30], [36]
Hamming	$H(U_i,V_i) = \begin{cases} 1, & \text{if } U_i = V_i \\ 0, & \text{if } U_i \neq V_i \end{cases}$ $Hamming(U,V) = \sum_{i=0}^{n-1} H(U_i,V_i)$	[33], [11]
Euclidean distance	$Euc(U,V) = \sqrt{\sum_{i=0}^{n-1}(U_i-V_j)^2}$	[24], [34]
Bhattacharya coefficient	$Bha = -ln(\sum_{x\in X}\sqrt{U(x)\cdot V(x)})$	[10]
Jaccard coefficient	$Jacc(U,V) = \frac{U^T V}{\parallel U-\bar{U} \parallel \cdot \parallel V-\bar{V} \parallel}$	[14], [17]

REFERENCES

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: Structure and dynamics, Physics reports, vol. 424, no. 4-5, pp. 175308, 2006.
[2] N. Mej, Networks: an introduction, 2010.
[3] S. Fortunato, Community detection in graphs, Physics reports, vol.486, no. 3-5, pp. 75174, 2010.
[4] S. Fortunato and D. Hric, Community detection in networks: A user guide, Physics Reports, vol. 659, pp. 144, 2016.
[5] R. J. Mokken, Cliques, clubs and clans, Quality and quantity, vol. 13, no. 2, pp. 161173, 1979.
[6] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature, vol. 435, no. 7043, p. 814, 2005.
[7] R. D. Alba, A graph-theoretic definition of a sociometric clique, Journal of Mathematical Sociology, vol. 3, no. 1, pp. 113126, 1973.
[8] Y. Cai, H.-f. Leung, Q. Li, H. Min, J. Tang, and J. Li, Typicality based collaborative filtering recommendation, IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 3, pp. 766779, 2014.
[9] X.-L. Zheng, C.-C. Chen, J.-L. Hung, W. He, F.-X. Hong, and Z. Lin,A hybrid trust-based recommender system for online communities of practice, IEEE Transactions on Learning Technologies, no. 1, pp. 11, 2015.

[10] B. K. Patra, R. Launonen, V. Ollikainen, and S. Nandi, A new similarity measure using bhattacharyya coefficient for collaborative filtering in sparse data, Knowledge-Based Systems, vol. 82, pp. 163177, 2015.
[11] M.-H. Chen, C.-H. Teng, and P.-C. Chang, Applying artificial immune systems to collaborative filtering for movie recommendation, Advanced Engineering Informatics, vol. 29, no. 4, pp. 830839, 2015.
[12] C. Yu and L. Huang, A web service qos prediction approach based on time-and location-aware collaborative filtering, Service Oriented Computing and Applications, vol. 10, no. 2, pp. 135149, 2016.
[13] A. Hernando, J. Bobadilla, and F. Ortega, A non negative matrix factorization for collaborative filtering recommender systems based on a bayesian probabilistic model, Knowledge-Based Systems, vol. 97, pp. 188202, 2016.
[14] W.-P. Lee and C.-Y. Ma, Enhancing collaborative recommendation performance by combining user preference and trust-distrust propagation in social networks, Knowledge-Based Systems, vol. 106, pp. 125134, 2016.
[15] N. Polatidis and C. K. Georgiadis, A dynamic multi-level collaborative filtering method for improved recommendations, Computer Standards & Interfaces, vol. 51, pp. 1421, 2017.
[16] M. M. Azadjalal, P. Moradi, A. Abdollahpouri, and M. Jalili, A trust-aware recommendation method based on pareto dominance and confidence concepts, Knowledge-Based Systems, vol. 116, pp. 130143, 2017.
[17] S. KG and G. S. Sadasivam, Modified heuristic similarity measure for personalization using collaborative filtering technique, Appl. Math, vol. 11, no. 1, pp. 307315, 2017.
[18] P. Liu, X. Wang, X. Che, Z. Chen, and Y. Gu, Defense against sybil attacks in directed social networks, in 2014 19th International Conference on Digital Signal Processing, 2014.
[19] G. Stringhini, P. Mourlanne, G. Jacob, M. Egele, C. Kruegel, and G. Vigna, EVILCOHORT: Detecting communities of malicious accounts on online services, in 24th USENIX Security Symposium (USENIX Security 15). Washington, D.C.: USENIX Association, pp. 563578, 2015.
[20] Y. Boshmaf, D. Logothetis, G. Siganos, J. Lería, J. Lorenzo, M. Ripeanu, K. Beznosov, and H. Halawa, Íntegro: Leveraging victim prediction for robust fake account detection in large scale osns, Computers & Security, vol. 61, pp. 142168, 2016.
[21] S. Misra, A. S. M. Tayeen, and W. Xu, Sybilexposer: An effective scheme to detect sybil communities in online social networks, in 2016 IEEE International Conference on Communications (ICC), pp.16, 2016.
[22] D. Mulamba, I. Ray, and I. Ray, Sybilradar: A graph-structure based framework for sybil detection in on-line social networks, in IFIP International Information Security and Privacy Conference. Springer, pp. 179193, 2016.
[23] J. Jia, B. Wang, and N. Z. Gong, Random walk based fake account detection in online social networks, in Dependable Systems and Networks (DSN), 2017 47th Annual IEEE/IFIP International Conference on. IEEE, pp. 273284, 2017.
[24] A. Klein, F. Ishikawa, and S. Honiden, Towards network-aware service composition in the cloud, in Proceedings of the 21st international conference on World Wide Web. ACM, 2012, pp. 959968.
[25] C. Cherifi, Y. Rivierre, and J.-F. Santucci, A community based algorithm for large scale web service composition, arXiv preprint arXiv:1305.0187, 2013.
[26] T. Li and T. He, Privacy-aware web services selection and composition,in Service Sciences (ICSS), 2014 International Conference on. IEEE, 2014, pp. 147151.
[27] S. Chhun, K. Malang, C. Cherifi, N. Moalla, and Y. Ouzrout, A web service composition framework based on centrality and community structure, in Signal-Image Technology & Internet-Based Systems (SITIS), 2015 11th International Conference on. IEEE, 2015, pp. 489496.
[28] F. Li, L. Zhang, Y. Liu, Y. Laili, and F. Tao, A clustering network-based approach to service composition in cloud manufacturing, International Journal of Computer Integrated Manufacturing, vol. 30, no. 12, pp. 13311342, 2017.
[29] W. Pan and C. Chai, Structure-aware mashup service clustering for cloud-based internet of things using genetic algorithm based clustering algorithm, Future Generation Computer Systems, 2018.
[30] C. Cobos, H. Munoz-Collazos, R. Urbano-Munoz, M. Mendoza, E. Leon, and E. Herrera-Viedma, Clustering of web search results based on the cuckoo search algorithm and balanced bayesian information criterion, Information Sciences, vol. 281, pp. 248264, 2014.
[31] X. He, M.-Y. Kan, P. Xie, and X. Chen, Comment-based multi-view clustering of web 2.0 items, in Proceedings of the 23rd international conference on World wide web. ACM, 2014, pp. 771782.
[32] C. H. Tseng, Y. H. Chen, C. C. Chuang, J. H. Wu, Y. S. Yang, and Y. W. Liang, Keen-means: A web page clustering tool based on an selfadjustable k-means algorithm, in Ubi-Media Computing and Workshops (UMEDIA), 2014 7th International Conference on. IEEE, 2014, pp. 300304.
[33] R. Forsati, A. Moayedikia, and M. Shamsfard, An effective web page recommender using binary data clustering, Information Retrieval Journal, vol. 18, no. 3, pp. 167214, 2015.
[34] L. M. Abualigah, A. T. Khader, M. A. Al-Betar, and M. A. Awadallah, A krill herd algorithm for efficient text documents clustering, in Computer Applications & Industrial Electronics (ISCAIE), 2016 IEEE Symposium on. IEEE, 2016, pp. 6772.
[35] R. Katarya and O. P. Verma, An effective web page recommender system with fuzzy c-mean clustering, Multimedia Tools and Applications, vol. 76, no. 20, pp. 21 48121 496, 2017.
[36] Y. Yan, G. Liu, S. Wang, J. Zhang, and K. Zheng, Graph-based clustering and ranking for diversified image search, Multimedia Systems, vol. 23, no. 1, pp. 4152, 2017.
[37] X. Yang, Y. Guo, Y. Liu, and H. Steck, A survey of collaborative filtering based social recommender systems, Computer Communications, vol. 41, pp. 110, 2014.
[38] L. Xiaojun, An improved clustering-based collaborative filtering recommendation algorithm, Cluster Computing, vol. 20, no. 2, pp. 12811288, 2017.
[39] X. Ma, H. Lu, Z. Gan, and J. Zeng, An explicit trust and distrust clustering based collaborative filtering recommendation approach, Electronic Commerce Research and Applications, vol. 25, pp. 2939, 2017.
[40] S. Kant and T. Mahara, Merging user and item based collaborative filtering to alleviate data sparsity, International Journal of System Assurance Engineering and Management, vol. 9, no. 1, pp. 173179, 2018.
[41] R. Hu, W. Dou, and J. Liu, Clubcf: a clustering-based collaborative filtering approach for big data application, IEEE transactions on emerging topics in computing, vol. 2, no. 3, pp. 302313, 2014.
[42] K. K. Fletcher and X. F. Liu, A collaborative filtering method for personalized preference-based service recommendation, in Web Services (ICWS), 2015 IEEE International Conference on. IEEE, 2015, pp. 400407.
[43] W. Wei, F. Xu, C. C. Tan, and Q. Li, Sybildefender: Defend against sybil attacks in large social networks, in INFOCOM, 2012 Proceedings IEEE. IEEE, 2012, pp. 19511959.
[44] Z. Chen, W. Hendrix, and N. F. Samatova, Community-based anomaly detection in evolutionary networks, Journal of Intelligent Information Systems, vol. 39, no. 1, pp. 5985, 2012.
[45] S. Nilizadeh, F. Labr'eche, A. Sedighian, A. Zand, J. Fernandez, C. Kruegel, G. Stringhini, and G. Vigna, Poised: Spotting twitter spam off the beaten paths, in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017, pp. 11591174.
[46] P. H. Meland, J. B. Guerenabarrena, and D. Llewellyn-Jones, The challenges of secure and trustworthy service composition in the future internet. in SoSE, 2011, pp. 329334.
[47] J. Shang, L. Liu, and C. Wu, Wscn: Web service composition based on complex networks, in Service Sciences (ICSS), 2013 International Conference on. IEEE, 2013, pp. 208213.
[48] M. Toyoda and M. Kitsuregawa, Extracting evolution of web communities from a series of web archives, in Proceedings of the fourteenth ACM conference on Hypertext and hypermedia. ACM, 2003, pp. 2837.
[49] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, Extracting large-scale knowledge bases from the web, in VLDB, vol. 99, 1999, pp. 639650.
[50] D. S. Sisodia, S. Verma, and O. P. Vyas, Augmented intuitive dissimilarity metric for clustering of web user sessions, Journal of Information Science, vol. 43, no. 4, pp. 480491, 2017.
[51] K. Bharat and M. R. Henzinger, Improved algorithms for topic distillation in a hyperlinked environment, in Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 1998, pp. 104111.
[52] A. Di Marco and R. Navigli, Clustering and diversifying web search results with graph-based word sense induction, Computational Linguistics, vol. 39, no. 3, pp. 709754, 2013.