Future Generation Computer Systems 114 (2021) 290–293

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Editorial

New research methods & algorithms in social network analysis

David Camacho [a,*], M[a] Victoria Luzón [b], Erik Cambria [c]

[a] Departamento de Sistemas Informáticos, Universidad Politécnica de Madrid, Calle de Alan Turing, s/n, 28031 Madrid, Spain
[b] ETS Ingenierías Informática y de Telecomunicación, Universidad de Granada, C/Periodista Daniel Saucedo Aranda, s/n, E-18071 Granada, Spain
[c] School of Computer Science and Engineering, Nanyang Technological University, 50 Nanyang Ave, 639798, Singapore

ARTICLE INFO

Article history:
Available online 14 August 2020

Keywords:
Social network analysis
Community detection
Natural language processing
Sentiment analysis
Social media mining

ABSTRACT

The exponential growth of social media and online social networks (e.g., Facebook, Twitter, Instagram, and TikTok) has changed the daily lives of millions of people. The ease to accessing, gathering and processing available data and the high societal and industrial interest in such data have attracted the interest of a large of research disciplines. This special issue has been focused mainly on Data Science and Artificial Intelligence techniques, and their application to social network analysis. The issue provides a total of 12 selected papers (out of 65) that represent latest advances and developments in these areas.

1. Introduction

Social media have changed the way people communicate. Millions of people all over the world leverage social media to create, share, and discuss information and for networking. Anyone with an Internet connection can share their own experiences via a video, provide their personal opinion via a tweet, or show the world something interesting via a picture. This has led social media analytics (SMA) to an exponential growth over the last years due to the amount of data continuously shared on the Internet [1,2]. Gohfar Khan defines SMA as "*the art and science of extracting valuable hidden insights from vast amounts of semistructured an unstructured social media data to enable informed and insightful decision making*" [3].

Different techniques have been created for analyzing opinions towards a product, predicting elections results, studying how fake news spread through social networks. The increasing interest in the aggregation and the analysis of social media information has given rise to growing interest in various areas of research, not only those related to computer science, but in others such as physics, psychology, marketing, tourism, and finance among others [4–7]. Specific topics include clustering, graph mining, community detection, natural language processing, entity disambiguation, information fusion, sentiment analysis, or recommender systems [8–11], to mention only few. Some of the current challenges in the area of social network analysis (SNA) involve big data analytics (data gathering, pre-processing, etc.), information fusion, scalability, online and streaming SNA systems, statistical modeling for large networks, pattern modeling and extraction, or visualization [12].

Therefore, this special issue focuses on:

1. the application of advanced data science and artificial intelligence techniques for knowledge extraction from social networks, and
2. the application of machine learning, soft computing, and computational intelligence to complex social media-based domains.

For this special issue, we received 65 articles, of which 12 were carefully selected. Next section shows a brief description of the main contents of each article. These works show some of the latest results and advances in the application of SNA in a wide range of domains and data sources.

2. Contents of the issue

The papers selected for this special issue reflect some of the current trends in SMA [13], SNA [14–16], information theory [17], natural language processing [18,19], sentiment analysis and opinion mining [18,20], community detection [14,21], machine learning (clustering, support vector machines, CNN, deep learning, RNN, etc.) [13,18,20,22], and evolutionary computation & metaheuristics [14,16,23,24]. The main findings and contributions of each paper are listed below.

- Atzeni et al. [18] present a state-of-the-art approach for sentiment polarity classification. This approach relies on an ensemble of bidirectional long short-term memory (LSTM)

* Corresponding author.
E-mail addresses: david.camacho@upm.es (D. Camacho), luzon@ugr.es (M.V. Luzón), cambria@ntu.edu.sg (E. Cambria).
URLs: http://aida.etsisi.upm.es/ (D. Camacho), https://lsi.ugr.es/lsi/luzon (M.V. Luzón), https://sentic.net/erikcambria/ (E. Cambria).

https://doi.org/10.1016/j.future.2020.08.006
0167-739X/

networks equipped with a neural attention mechanism. The system makes use of pre-trained word embeddings, and is capable of predicting new vectors for out-of-vocabulary words, by learning distributional representations based on word spellings. During the training process the recurrent neural network is used to perform a fine-tuning of the original word embeddings, taking into account information about sentiment polarity. This step can be particularly helpful for sentiment analysis, as word embeddings are usually built based on context information, while words with opposite sentiment polarity often occur in similar contexts. The system described in this paper is an improved version of an approach that competed in a recent challenge on semantic sentiment analysis. The performance of the system is tested using the data from this challenge, showing that this approach allows reaching better results with respect to the previous top-scoring system. Finally, the authors embedded the proposed sentiment polarity approach on top of a humanoid robot to identify the sentiment of speakers in real time.

- Benke et al. [22] approached the problem of topic classification in social networks datasets collected from Twitter and Colab. The paper proposes to apply machine learning algorithms to the urban data generated by social networks in order to create classifiers to automatically categorize citizens messages according to the different cities services dimensions. The main question studied by authors is how to determine a supervised learning model to classify messages into Smart City dimensions using ISO 37120 as the standard model. For that, a wrapper approach was used considering five methods of features selection that were tested through two nominated machine learning algorithms (LSVC and LR for Colab, and CNB and DTC for Twitter) among eight tested for each dataset. Two distinct text datasets in Portuguese were collected from two social networks: Twitter (1,950 tweets) and Colab.re (65,066 posts).
- In Ma et al. [21] a new method is proposed to identify the most influential nodes which are considered as cores of communities and achieve the initial communities. Then, by an expansion strategy, unassigned nodes are added to initial communities to expand communities. Finally, merging overlapping communities to get the final community structure. Authors propose a new centrality measure based on local and global information (LGI), and a community detection algorithm based on seed expansion (LGIEM). The proposed centrality LGI combines local information with global information of nodes. If a node has high influence, the node can be regarded as a seed or core node. In this new method, these influential nodes and their neighbor constitute initial communities. To evaluate the performance of the proposed node influence method (LGI), the susceptible–infected–removed (SIR) diffusion model are used. Testing with both synthetic networks and real-world networks shows that LGI can identify best nodes with high influence and is better than other centrality methods. Finally, experiments show that the proposed community detection algorithm based on influential nodes (LGIEM) is able to detect communities efficiently, and achieves better performance compared to other recent methods.
- The main idea presented by Loia & Orciuoli [15] is to use the historical data provided by global terrorism database (GTD), a dataset stemming from the START project which offers information related to terrorist attacks perpetrated since 1970. GTD can be analyzed in order to provide, for instance, prediction models capable to conceptualize the behaviors of terrorist groups in specific time intervals. Such conceptualizations are, subsequently, used to understand the similarity between terrorist groups and elicit relations to represent terrorists' networks. The above networks can be used to study the temporal evolution of terrorist groups' behaviors by applying the approach in different time periods and studying differences among the resulting networks. The approach is mainly based on rough set theory and three-way decisions theory and provides an original similarity function based on the definition of boundary regions.
- The paper by Pandya et al. [13] presents the problem of age prediction from a Twitter dataset, where the prediction issue is viewed as a classification task. For this purpose, an innovative model based on a convolutional neural network (CNN) is devised. To this end, the authors rely on language-related features and social media specific metadata. More specifically, two features that have not been previously considered in the literature are introduced: the content of URLs and hashtags appearing in tweets. They also employ distributed representations of words and phrases present in tweets, hashtags and URLs, pre-trained on appropriate corpora in order to exploit their semantic information in age prediction. Finally, the experimental findings show that this CNN-based classifier, when compared with baseline models, yields an improvement of up to 12.3% for Dutch dataset, 9.8% for English1 dataset, and 6.6% for English2 dataset in the micro-averaged F1 score.
- The work presented by Panizo et at. [14] is focused on the design, implementation, and the empirical analysis of a new multi-objective genetic algorithm that combines an Immigrant's scheme with local search strategies for dynamic community detection. The main contribution of this new algorithm is to address the adaptation of these strategies to dynamic networks. On the one hand, the immigrant's scheme motif is to reuse previously acquired information to reduce computational time. On the other hand, in a dynamic environment is possible that a valid solution became invalid due to some changes in the environment, for example, if some nodes or edges have been removed or added to the network. Therefore, the aim of the local search operator used in the new algorithm is to transform an invalid solution, due to a change happened on the network, into a valid one maintaining the highest possible quality. Finally, the proposed algorithm has been tested using several synthetic and real-world networks, and compared against several state-of-the-art algorithms (DYNMOGA, ALPA, Infomap).
- The work by Redondo et al. [17] combines two research areas, information theory (entropy analysis) and SMA (clustering techniques), to pursue an efficient mechanism able to detect anomalies in the dynamics of citizens in urban areas. The proposed approach introduces a hybrid solution based on previous techniques to early detect unexpected behaviors in the social media activity, which can be seen as a trustful evidence of unexpected changes of the activity in the city. Location-based Social Networks (LBSNs) provide an interesting source of geo-located data that have been previously used to obtain patterns of the dynamics of crowds throughout urban areas, the authors show how activity in LBSNs reflects the real activity in the city, through a set of experiments over a data set collected from Instagram for seven months in New York City.
- The work by Sanchez-Oro & Duarte [16] presents a Multi-start Iterated Greedy algorithm (MSIG) for detecting communities in social networks. In particular, it is focused on the optimization of the modularity value, which is a metric traditionally used in unsupervised community detection, where the optimal solution is not known a priori. The MSIG

method is conformed by a new procedure for generating initial solutions that follows a greedy criterion maximizing the modularity value. Additionally, new destructive and constructive phases for escaping from local optima are proposed to increase the diversification component of the algorithm, resulting in a wider exploration of the search space. The results obtained have been compared against the best previous methods found in the literature, emerging MSIG as the best performing procedure.

- In Shi et al. [19], the authors address the problem of timely detection of meteorological events with the aim of early warning information. They overcome the traditional meteorological event detection methods, which rely on physical sensors, which is usually costly and inflexible, presenting a new model, as a new form of lightweight social sensor. A sentence-level feature-based meteorological event detection model to detect 14 types of meteorological events defined by the China Meteorological Administration (CMA) in Sina Weibo is presented. This new joint model consists of two modules: a fine-tuned BERT as the language model and a wide-grained capsule network as the event detection network. The design of the proposed model considers the correlation among meteorological events and achieves the best results on all metrics compared with other baseline models. Moreover, as a practical application, the model has been applied to the meteorological event monitoring platform in the CMA Public Meteorological Service Center to provide online services.
- The work developed by Varietti et al. [20] is focused on a highly relevant problem related to sentiment analysis, the study and analysis of the underlying structural information of opinions. To do that, social media sources are used to gather user opinions, which are structured in most of the cases in two sections: the title and the content or body of the opinion. Authors claim that the structure of social media opinions has useful information for the polarity classification task. To demonstrate this hypothesis, a new model for optimizing the contribution of underlying structural information for polarity classification is proposed. This model is built by weighting the contribution of each section, title and body. Authors develop a modified SVM that includes a weight parameter, which is optimized via a line-search strategy. The model is evaluated on three datasets of reviews from different domains written in two different variations of the Spanish language (Spanish spoken in Spain and in Chile). The results show that the model outperforms the classification of the joint or individual classification of each section of the opinion.
- Wang et al. [23] address a well-known (NP-hard) problem in social networks, the Influence Maximization (IM), where challenge is to determine the initial diffusion seed nodes especially when the size of social network increases. In this paper, firstly a new index, named node key degree (NKD), to denote the significance degree of each node is introduced. A node's NKD is determined by two factors: (1) the number of its direct previous nodes, and (2) the number of its successor off-springs within a certain number of levels. Then, we propose a novel efficient ITÖ Algorithm to solve the IM problem, termed as ITÖ-IM. In order to show the strengths and weaknesses of ITÖ-IM, an extensive computational studies on the six real world datasets have been carried out by authors. The experimental results show that the new algorithm achieves competitive results in influence spread as compared with other four state-of-the-art algorithms in the large-scale social networks.
- Finally, the work presented by Yao et al. [24] propose a greedy algorithm named RSEMP algorithm, which takes the Restricted Isometry Property (RIP) criterion as theoretical basis of signal sparsity estimation, which is used for determining the number of selection atoms. Using the relations between of the residuals to adjust the step size of the atom selection for pursuing the best matching atoms set to achieve a more excellent reconstruction performance. This new method tries to improve some the existing greedy algorithms whose reconstruction performance are effected by signal sparsity, and the other greedy algorithms whose reconstruction performance are not constrained by signal sparsity have the problems of unreasonable long reconstruction time, or low reconstruction precision. The proposed algorithm can flexibly remove the inappropriate atoms and gradually approach the sparsity adaptively when the step size is under control. This is the reason why RSEMP algorithm can approximate to the original signal with a high precision. Experimental results show that the ability of RSEMP algorithm to provide a better reconstruction quality and a less computational time under random measurement matrix and determinacy measurement matrix when the sparsity is estimated in advance.

3. Conclusion

The fast growth of social media and online social networks, and their ease of access, makes these areas an excellent test bed for other popular research disciplines such as machine learning, natural language processing, graph computing, and evolutionary computation.

This issue has collected a small number of high-quality works in such disciplines, to show how new methods and algorithms are being applied to SNA research. The editors hope that this small collection of research works will be useful for FGCS readers.

CRediT authorship contribution statement

David Camacho: Writing, Special issue definition, Dissemination, Revision, Selection. **Mª Victoria Luzón:** Writing, Special issue definition, Dissemination, Revision, Selection. **Erik Cambria:** Writing, Special issue definition, Dissemination, Revision, Selection.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The guest editors of this special issue are especially grateful to the many anonymous reviewers who have contributed with their time and expertise in reviewing and improving the selected articles. Without their precious efforts, this special issue would not have been possible.

The guest editors would also like to thank the Editor-in-Chief, Prof. Peter Sloot, for his help and support throughout all stages of the special issue.

References

[1] G. Bello-Orgaz, J.J. Jung, D. Camacho, Social big data: Recent achievements and new challenges, Inf. Fusion 28 (2016) 45–59.
[2] A. Hussain, E. Cambria, Semi-supervised learning for big social data analysis, Neurocomputing 275 (2018) 1662–1673.
[3] G.F. Khan, Social media analytics, in: Social Media for Government, Springer, 2017, pp. 93–118.
[4] E. Cambria, M. Grassi, A. Hussain, C. Havasi, Sentic computing for social media marketing, Multimedia Tools Appl. 59 (2) (2012) 557–577.
[5] F. Xing, E. Cambria, R. Welsch, Natural language based financial forecasting: A survey, Artif. Intell. Rev. 50 (1) (2018) 49–73.
[6] G. Bello-Orgaz, R.M. Mesas, C. Zarco, V. Rodriguez, O. Cordón, D. Camacho, Marketing analysis of wineries using social collective behavior from users' temporal activity on twitter, Inf. Process. Manage. (2020) 102220.
[7] A. Valdivia, E. Hrabova, I. Chaturvedi, M.V. Luzón, L. Troiano, E. Cambria, F. Herrera, Inconsistencies on tripadvisor reviews: A unified index between users and sentiment analysis methods, Neurocomputing 353 (2019) 3–16.
[8] S. Cavallari, E. Cambria, H. Cai, K. Chang, V. Zheng, Embedding both finite and infinite communities on graph, IEEE Comput. Intell. Mag. 14 (3) (2019) 39–50.
[9] Y. Li, S. Wang, Q. Pan, H. Peng, T. Yang, E. Cambria, Learning binary codes with neural collaborative filtering for efficient recommendation systems, Knowl.-Based Syst. 172 (2019) 64–75.
[10] G. Bello-Orgaz, S. Salcedo-Sanz, D. Camacho, A multi-objective genetic algorithm for overlapping community detection based on edge encoding, Inform. Sci. 462 (2018) 290–314.
[11] A. Valdivia, M. Luzón, F. Herrera, Sentiment analysis in tripadvisor, IEEE Intell. Syst. 32 (4) (2017) 72–77.
[12] D. Camacho, A. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, E. Cambria, The four dimensions of social network analysis: An overview of research methods, applications, and software tools, Inf. Fusion 63 (2020) 88–120.
[13] A. Pandya, M. Oussalah, P. Monachesi, P. Kostakos, On the use of distributed semantics of tweet metadata for user age prediction, Future Gener. Comput. Syst. 102 (2020) 437–452.
[14] A. Panizo-L.L.edot, G. Bello-Orgaz, D. Camacho, A multi-objective genetic algorithm for detecting dynamic communities using a local search driven immigrant's scheme, Future Gener. Comput. Syst. (2019).
[15] V. Loia, F. Orciuoli, Understanding the composition and evolution of terrorist group networks: A rough set approach, Future Gener. Comput. Syst. 101 (2019) 983–992.
[16] J. Sánchez-Oro, A. Duarte, Iterated greedy algorithm for performing community detection in social networks, Future Gener. Comput. Syst. 88 (2018) 785–791.
[17] R.P.D. Redondo, C. Garcia-Rubio, A.F. Vilas, C. Campo, A. Rodriguez-Carrion, A hybrid analysis of LBSN data to early detect anomalies in crowd dynamics, Future Gener. Comput. Syst. 109 (2020) 83–94.
[18] M. Atzeni, D.R. Recupero, Multi-domain sentiment analysis with mimicked and polarized word embeddings for human–robot interaction, Future Gener. Comput. Syst. (2019).
[19] K. Shi, C. Gong, H. Lu, Y. Zhu, Z. Niu, Wide-grained capsule network with sentence-level feature to detect meteorological event in social network, Future Gener. Comput. Syst. 102 (2020) 323–332.
[20] C. Vairetti, E. Martínez-Cámara, S. Maldonado, V. Luzón, F. Herrera, Enhancing the classification of social media opinions by optimizing the structural information, Future Gener. Comput. Syst. 102 (2020) 838–846.
[21] T. Ma, Q. Liu, J. Cao, Y. Tian, A. Al-Dhelaan, M. Al-Rodhaan, LGIEM: Global and local node influence based community detection, Future Gener. Comput. Syst. 105 (2020) 533–546.
[22] L. Bencke, C. Cechinel, R. Munoz, Automated classification of social network messages into smart cities dimensions, Future Gener. Comput. Syst. (2020).
[23] Y. Wang, W. Dong, X. Dong, A novel ITÖ Algorithm for influence maximization in the large-scale social networks, Future Gener. Comput. Syst. 88 (2018) 755–763.
[24] S. Yao, A.K. Sangaiah, Z. Zheng, T. Wang, Sparsity estimation matching pursuit algorithm based on restricted isometry property for signal reconstruction, Future Gener. Comput. Syst. 88 (2018) 747–754.

David Camacho is currently working as Associate Professor with the Departamento de Sistemas Informáticos at Universidad Politécnica de Madrid (Spain) and leads the Applied Intelligence and Data Analysis group (AIDA). He received a Ph.D. in Computer Science (2001) from Universidad Carlos III de Madrid, and a B.S. in Physics (1994) from Universidad Complutense de Madrid. His research interests include Data Mining, Evolutionary Computation, Social Network Analysis, and Swarm Intelligence, among others. He is on the editorial board of several journals, such as Information Fusion, International Journal of Bioinspired Computation, and Expert systems. Contact him at: david.camacho@upm.es.

M[a] Victoria Luzón is an Associate Professor in the Software Engineering Department at University of Granada, Spain. Her research interests include NLP, sentiment analysis, artificial intelligence, and federated learning. Dr. Luzón has a Ph.D. in industrial engineering from the University of Vigo. Contact her at: luzon@ugr.es.

Erik Cambria is an Associate Professor at NTU, where he also holds the appointment of Provost Chair in Computer Science and Engineering. Prior to joining NTU, he worked at Microsoft Research Asia and HP Labs India and earned his Ph.D. through a joint programme between the University of Stirling and MIT Media Lab. He is Associate Editor of several journals, e.g., INFFUS, IEEE CIM, and KBS, Department Editor of IEEE Intelligent Systems, Editorial Board Chair of Cognitive Computation, and is involved in many international conferences as area/program chair and invited speaker. Contact him at: cambria@ntu.edu.sg.