using supervised techniques and NLP by utilizing their social and linguistics behavior [17] Some of these works are empirical studies without developing a systematic methodology. Recent efforts include analyzing the dynamics of the black-market of hacking services [18], extracting malicious IP addresses reported by users in security forums [19]. A recent work [8], REST, identifies and classifies threads given keywords of interest, and we use it to validate our cluster labeling.

b. Mining social networks and other types of forums: Researchers have studied a wide range of online media such as blogs, commenting platforms, Reddit, Facebook etc. Some recent works analyze the user behavioral patterns observed in Reddit [20] and infer information for the users from their activities on Facebook [21] and GitHub [22], [23]. Despite some common algorithmic foundations, we argue that different media and different questions require novel and targeted methods.

Event detection is a broad and related type of research [11], [24]. A recent work [12] proposes a hierarchical multi-aspect attention approach for event detection but does not consider the author and temporal dimension as we do here.

c. Tensor Decomposition approaches: Tensors is a well-studied area with a wide range of diverse applications and domains [5] including understanding the multilingual social networks in online immigrant communities [25], community assignment of nodes in multi-aspect graph [26], and tensor-based community evolution [27]. We are not aware of any tensor-based event extraction studies for online forums. In our work, we adapted the CP tensor decomposition [5], [28] and combined it with L1 regularization to filter out the insignificant entities.

VI. Conclusion

We propose, TenFor, an unsupervised-learning tensor-based approach to systematically identify important events in a three-dimensional space: (a) users, (b) threads, and (c) time. Our approach has three main advantages: (a) it operates in an unsupervised way, though the user has ways to influence its focus, if so desired, (b) it provides visual and intuitive information, and (c) it identifies both the events of interest, and the entities of interest within the event, including threads, users, and time intervals.

Our work is a step towards an automated unsupervised capability, which can allow security analysts and researchers to shift through the wealth of information that exists in security forum and online forums in general.

VII. Acknowledgement

This work was supported by the UC Multicampus-National Lab Collaborative Research and Training (UCNLCRT) award #LFR18548554.

References

[1] N. Shah, D. Koutra, T. Zou, B. Gallagher, and C. Faloutsos, "Timecrunch: Interpretable dynamic graph summarization," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 1055–1064.

[2] T. Alzahrani and K. J. Horadam, "Community detection in bipartite networks: Algorithms and case studies," in *Complex systems and networks*. Springer, 2016, pp. 25–50.

[3] Online Forums, "Ethical Hacker, Hack This Site, Offensive Community, MPGH," https://www.ethicalhacker.net/, https://www.hackthissite.org/, http://offensivecommunity.net/, https://mpgh.net/.

[4] S. Pastrana, D. R. Thomas, A. Hutchings, and R. Clayton, "Crimebb: Enabling cybercrime research on underground forums at scale," in *WWW*, 2018, pp. 1845–1854.

[5] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[6] E. E. Papalexakis, "Automatic unsupervised tensor mining with quality assessment," in *SDM16*. SIAM, 2016, pp. 711–719.

[7] A. Sapienza, A. Bessi, and E. Ferrara, "Non-negative tensor factorization for human behavioral pattern mining in online games," *Information*, vol. 9, no. 3, p. 66, 2018, multidisciplinary Digital Publishing Institute.

[8] J. Gharibshah, E. E. Papalexakis, and M. Faloutsos, "REST: A thread embedding approach for identifying and classifying user-specified information in security forums." *ICWSM*, 2020.

[9] K. Sheridan, T. G. Puranik, E. Mangortey, O. J. Pinon-Fischer, M. Kirby, and D. N. Mavris, "An application of dbscan clustering for flight anomaly detection during the approach phase," in *AIAA Scitech 2020 Forum*, 2020, p. 1851.

[10] D. Sharma, B. Kumar, and S. Chand, "A survey on journey of topic modeling techniques from svd to deep learning," *IJMECS*, vol. 9, no. 7, p. 50, 2017, modern Education and Computer Science Press.

[11] L.-L. Shi, L. Liu, Y. Wu, L. Jiang, M. Kazim, H. Ali, and J. Panneerselvam, "Human-centric cyber social computing model for hot-event detection and propagation," *IEEE Transactions on CSS*, vol. 6, no. 5, pp. 1042–1050, 2019, iEEE.

[12] S. Mehta, M. R. Islam, H. Rangwala, and N. Ramakrishnan, "Event detection using hierarchical multi-aspect attention," in *WWW*, 2019, pp. 3079–3085.

[13] A. Claster, "News of hacking by vandathegod:," https://www.databreaches.net/dozens-of-government-websites-defaced-by-vandathegod-hacktivists/, March 2019, [accessed March-2020].

[14] J. L. Fleiss and J. Cohen, "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability," *Educational and psychological measurement*, vol. 33, no. 3, pp. 613–619, 1973.

[15] L. Blog, "Major ransomware events:," https://blog.logsign.com/10-worst-ransomware-attacks-in-the-last-five-years/, March 2018.

[16] C. Carter, "Romantic scamming in gaming forum," https://www.stuff.co.nz/auckland/106254141/romantic-scammers-preying-on-players-of-online-game-words-with-friends/.

[17] E. Marin, J. Shakarian, and P. Shakarian, "Mining key-hackers on darkweb forums," in *ICDIS*. IEEE, 2018, pp. 73–80.

[18] R. S. Portnoff, S. Afroz, G. Durrett, J. K. Kummerfeld, T. Berg-Kirkpatrick, D. McCoy, K. Levchenko, and V. Paxson, "Tools for automated analysis of cybercriminal markets," in *WWW*, 2017, p. 657.

[19] J. Gharibshah, E. E. Papalexakis, and M. Faloutsos, "RIPEx: Extracting malicious ip addresses from security forums using cross-forum learning," in *PAKDD*. Springer, 2018.

[20] S. Thukral, H. Meisheri, T. Kataria, A. Agarwal, I. Verma, A. Chatterjee, and L. Dey, "Analyzing behavioral trends in community driven discussion platforms like reddit," in *ASONAM*. IEEE, 2018, pp. 662–669.

[21] J. Bayer, N. Ellison, S. Schoenebeck, E. Brady, and E. B. Falk, "Facebook in context: Measuring emotional responses across time and space," *new media & society*, vol. 20, no. 3, pp. 1047–1067, 2018, sAGE.

[22] M. O. F. Rokon, R. Islam, A. Darki, E. E. Papalexakis, and M. Faloutsos, "Sourcefinder: Finding malware source-code from publicly available repositories in github," in *23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*. USENIX, 2020, pp. 149–163.

[23] R. Islam, M. O. F. Rokon, A. Darki, and M. Faloutsos, "Hackerscope: The dynamics of a massive hackeronline ecosystem," in *Proceedings of International Conference on Advances in Social Network Analysis and Mining (ASONAM)*. IEEE/ACM, 2020.

[24] M. Hasan, M. A. Orgun, and R. Schwitter, "Real-time event detection from the twitter data stream using the twitternews+ framework," *Information Processing & Management*, vol. 56, p. 1146, 2019, elsevier.

[25] E. Papalexakis and A. S. Doğruöz, "Understanding multilingual social networks in online immigrant communities," in *WWW*, 2015, p. 865.

[26] E. Gujral and E. E. Papalexakis, "Smacd: Semi-supervised multi-aspect community detection," in *ICDM*. SIAM, 2018, pp. 702–710.

[27] Y. Liu, G. Yan, J. Ye, and Z. Li, "Community evolution based on tensor decomposition," in *ICPCSEE*. Springer, 2019, pp. 62–75.

[28] N. K. M. Faber, R. Bro, and P. K. Hopke, "Recent developments in candecomp/parafac algorithms: a critical review," *Chemometrics and Intelligent Laboratory Sys*, vol. 65, no. 1, pp. 119–137, 2003, elsevier.