Peeking at Visualization Research on Information Diffusion

M. Usul[1] and A. Arleo[2,1]

[1]TU Wien, Austria
[2]Eindhoven University of Technology, The Netherlands

Abstract

Diffusion Processes are a widely researched topic of interest to different scientific domains. One of the most popular research directions is Information Diffusion, pertaining how information spreads over a tightly connected network. From the modeling perspective, many different approaches are known in the literature; however, in the visualization community, this still represents an under-investigated problem. In this work, we present a succinct overview of the current state-of-the-art in Visual Analytics techniques employed in representing and understanding diffusion processes happening over networks. We consider different application domains and introduce a taxonomy that categorizes and provides structure to our selection of papers, fostering further research in the field of Visual Analytics of Information Diffusion processes.

CCS Concepts

• ***Human-centered computing*** → ***Visualization; Visual analytics;***

1. Introduction

With the term *Diffusion Process* we refer to a category of research problems across many fields of research, including, e.g., the investigation of epidemic pathogen transmission [BPW*21], or the analysis of Information Diffusion (ID) in large social networks [LFWT18]. Specifically, we are interested in the diffusion of information over networks. A *Network* (or *Graph*) G consists of a set of nodes (or vertices) V and edges E connecting them. An ID process describes the propagation of information (e.g., pathogens, malware, fake news) across a network or population, under the assumptions of a diffusion *model* that encapsulates the high-level behavior of the real phenomenon. Many such models exist (see, e.g., the survey by Li et al. [LFWT18]). However, the use of Visual Analytics (VA) in the context of ID is still a largely under-investigated research direction. Even though there have been surveys on the use of VA in ID, they primarily concentrate on VA tools within the application context of social networks [CLY17].

Our Contribution. We broaden the scope beyond the limits of social networks and investigate the state-of-the-art in VA of ID processes across different research domains. We categorize the 26 collected contributions in a taxonomy based on their main task, highlighting opportunities for further research, focusing on domain-independent approaches.

2. Methodology

We conducted a thorough literature research with clearly defined inclusion criteria. We aim at ID related papers that present a visualization to explore the progression of the process or the results of a simulation. We include VA systems and exclude static visualizations that offer no or limited interaction and do not enable an in-depth exploration of data. We also consider approaches that use data that inherently describes diffusion behavior, such as reposting [YWL*14] or citation networks [HSS*20]. We also include epidemic models that do not necessarily require an underlying graph structure but rather describe the diffusion process as a set of differential equations [MLR*11]. This process resulted in a total of 26 selected contributions.

3. The Taxonomy

Since we have collected approaches from different application domains, the categorization has to be suitable for this diversity and offer a degree of generalization. For this reason, we decided to focus on the VA aspect and the main goal that it aims to achieve. We look at what kind of tools are offered and to what extent the user can interact with the diffusion process in question to reach a specific goal. Our taxonomy in Table 1 distinguishes alongside three main criteria: (i) The visualization method, (ii) the main goal that is achieved by the VA approach, (iii) and a set of additional utility features provided by the VA system. In an iterative process, we extracted the following six main goals:

A - Diffusion Exploration Twelve of our collected approaches offer an extensive toolkit for the exploration of diffusion networks to unfold patterns, analyze trends, and detect key influencers. Methods that are presented in this section focus on the overall exploration of diffusion processes via a diverse toolkit and provide a comprehensive, often coordinated dashboard to assist in that goal.

DOI: 10.2312/evp.20241089

A variety of different visualization methods have been tried in this context. Baumgartl et al. [BPW*21] for example, leverage storyline visualization to explore epidemic outbreak data.

B - Spread Simulation Besides the general exploration of diffusion processes, some approaches focus on the simulation and prediction of diffusion processes. Eight of our collected papers fall into this classification. Half of them specifically apply to the domain of epidemic and pandemic response, three to social networks, and one to the financial domain. We identify Node-Link and Map-based visualization to be the main methods. Map-based methods are especially popular in the domain of epidemiology.

C - Influence Maximization (IM) IM is an optimization problem that, given a network and a diffusion model, aims at finding the smallest set of initial seed nodes that yields that largest spread [LFWT18]. It is deeply connected to (**B**) as it requires a model assumption (and simulations are required) but papers in this category have a specific interest for the algorithmic aspect of IM and ID, also due to its NP-Hardness. Arleo et al. [ADL*22] and Long and Wong [LW14] leverage VA and enable an in-depth exploration of IM models.

D - Influence Summarization For the summarization of time-evolving influence graphs, Huang et al. [HSS*20] propose new edge summarization algorithms for nodes, edges, and the temporal dimension to visualize the evolution of citation influence networks with a tree-like flow map.

E - Model Comparison Vallet et al. [VKPM15] lay their focus on the comparison of different diffusion models where the user can formally describe diffusion models through a set of graph rewriting rules and application strategies.

F - Anomaly Detection Zhao et al. [ZCW*14] use machine learning algorithms to extract anomalies from a Twitter dataset. The challenge and main focus here lie in distinguishing unconventional patterns, like the dissemination of rumors or misinformation (anomalies), from more traditional trends such as popular topics.

4. Takeaways and Conclusions

We presented a compact and structured overview of visualization research on ID by categorizing contributions on their visualization technique(s), main goal, and secondary tasks. From Table 1, it is possible to see that the main visualization techniques employed depend largely on the domain context (i.e., concerning the use of maps). Node-link is also the most employed visualization technique, being intuitive and easy to comprehend. However, it can quickly get cluttered, often requiring other accompanying visualization techniques to scale up [ADL*22]. Timelines are also frequently used, due to the dynamic nature of diffusion phenomena. Concerning tasks, (**A**) and (**B**) are the ones presenting the largest amount of papers, suggesting more numerous un-investigated potential research directions for the remaining categories (**C**-**F**). Considering additional tasks, comparison is the least investigated. This work is intended as a starting point toward a formal characterization of the ID problem in visualization and to raise awareness of the more and less researched topics in this domain. A first step in this direction could be the creation of a task taxonomy. This would support research, development, and comparison of both general and domain-oriented VA approaches that could be used to analyze and explore complex ID processes in multiple application domains.

Approaches	Node-Link	Matrix	Map Metaphor	Map	River Metaphor	Tree	Timeline	Storyline	Main Goal	Key-Player Detection	Comparison Tools	Parameter Tweaking
[NYX*12]	■			■			□		A	■		
[STP*17]	■						□		A		■	
[HNWC23]		■	□				□		A	■	■	■
[CCL*17]			■				□		A	■		
[CCW*19]			■			□	□		A	■		
[CLCY20]			■				□		A	■		
[MBB*11]				■			■		A			
[WLY*14]	□				■	□	■		A	■	■	
[SWL*14]					■		■		A	■	■	
[DXZ*14]				□		■	□		A	■		
[BPW*21]	□						□	■	A	■		
[YWL*14]	■					■			A	■	■	
[VLDBF15]	■								B		■	■
[SRMV16]	■								B	■		■
[ST20]	■								B			■
[MLR*11]				■					B			■
[BGG*11]				■					B			■
[YDH*17]				■					B		■	■
[AME11]				■		■	■		B		■	■
[LWY*20]					■	■	■		B			■
[ADL*22]	■	■							C	■	■	■
[LW14]				■					C	■		■
[HSS*20]						■	□		D	■		■
[STTL15]	■								D	■		■
[VKPM15]	■								E		■	■
[ZCW*14]					■	□			F	■		■

Table 1: *Taxonomy of all 26 contributions classified as in Section 3. Pink cells refer to visualization methods that serve an auxiliary or navigational purpose, red ones are used in main views.*

Acknowledgements. This research was conducted within the Austrian Science Fund (FWF) grant SANE [10.55776/I6635]. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

References

[ADL*22] Arleo A., Didimo W., Liotta G., Miksch S., Montecchiani F.: Influence Maximization With Visual Analytics. *IEEE Trans. on Vis. and Computer Graphics 28*, 10 (Oct. 2022), 3428–3440. doi:10.1109/TVCG.2022.3190623. 2

[AME11] Afzal S., Maciejewski R., Ebert D. S.: Visual analytics decision support environment for epidemic modeling and response evaluation. In *2011 IEEE Conf. on Visual Analytics Science and Technology (VAST)* (Providence, RI, USA, Oct. 2011), IEEE, pp. 191–200. doi:10.1109/VAST.2011.6102457. 2

[BGG*11] Broeck W. V. D., Gioannini C., Gonçalves B., Quaggiotto M., Colizza V., Vespignani A.: The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infectious Diseases 11*, 1 (Dec. 2011), 37. doi:10.1186/1471-2334-11-37. 2

[BPW*21] Baumgartl T., Petzold M., Wunderlich M., Höhn M., Archambault D., Lieser M., Dalpke A., Scheithauer S., Marschollek M., Eichel V. M., Mutters N. T., Consortium H., von Landesberger T.: In Search of Patient Zero: Visual Analytics of Pathogen Transmission Pathways in Hospitals. *IEEE Trans. on Vis. and Computer Graphics 27*, 2 (Feb. 2021), 711–721. arXiv:2008.09552 [cs]. doi:10.1109/TVCG.2020.3030437. 1, 2

[CCL*17] Chen S., Chen S., Lin L., Yuan X., Liang J., Zhang X.: E-Map: A Visual Analytics Approach for Exploring Significant Event Evolutions in Social Media. In *2017 IEEE Conf. on Visual Analytics Science and Technology (VAST)* (Phoenix, AZ, Oct. 2017), IEEE, pp. 36–47. doi:10.1109/VAST.2017.8585638. 2

[CCW*19] Chen S., Chen S., Wang Z., Liang J., Wu Y., Yuan X.: D-Map+: Interactive Visual Analysis and Exploration of Ego-centric and Event-centric Information Diffusion Patterns in Social Media. *ACM Trans. on Intelligent Systems and Technology 10*, 1 (Jan. 2019), 1–26. doi:10.1145/3183347. 2

[CLCY20] Chen S., Li S., Chen S., Yuan X.: R-Map: A Map Metaphor for Visualizing Information Reposting Process in Social Media. *IEEE Trans. on Vis. and Computer Graphics 26*, 1 (Jan. 2020), 1204–1214. doi:10.1109/TVCG.2019.2934263. 2

[CLY17] Chen S., Lin L., Yuan X.: Social Media Visual Analytics. *Computer Graphics Forum 36*, 3 (June 2017), 563–587. doi:10.1111/cgf.13211. 1

[DXZ*14] Donghao Ren, Xin Zhang, Zhenhuang Wang, Jing Li, Xiaoru Yuan: WeiboEvents: A Crowd Sourcing Weibo Visual Analytic System. In *2014 IEEE Pacific Vis. Symposium* (Yokohama, Mar. 2014), IEEE, pp. 330–334. doi:10.1109/PacificVis.2014.38. 2

[HNWC23] Han D., Nayeem A.-A.-R., Windett J., Cho I.: PDViz: A Visual Analytics Approach for State Policy Data. *Computer Graphics Forum 42*, 1 (Feb. 2023), 305–318. Publisher: John Wiley & Sons, Ltd. doi:10.1111/cgf.14732. 2

[HSS*20] Huang Y., Shi L., Su Y., Hu Y., Tong H., Wang C., Yang T., Wang D., Liang S.: Eiffel: Evolutionary Flow Map for Influence Graph Vis. *IEEE Trans. on Vis. and Computer Graphics 26*, 10 (Oct. 2020), 2944–2960. doi:10.1109/TVCG.2019.2906900. 1, 2

[LFWT18] Li Y., Fan J., Wang Y., Tan K.-L.: Influence maximization on social graphs: A survey. *IEEE Trans. on Knowledge and Data Engineering 30*, 10 (2018), 1852–1872. doi:10.1109/TKDE.2018.2807843. 1, 2

[LW14] Long C., Wong R. C.-W.: Visual-VM: A Social Network Vis. Tool for Viral Marketing. In *2014 IEEE Intl. Conf. on Data Mining Workshop* (Shenzhen, China, Dec. 2014), IEEE, pp. 1223–1226. doi:10.1109/ICDMW.2014.40. 2

[LWY*20] Li Q., Wu Z., Yi L., N. K. S., Qu H., Ma X.: WeSeer: Visual Analysis for Better Information Cascade Prediction of WeChat Articles. *IEEE Trans. on Vis. and Computer Graphics 26*, 2 (Feb. 2020), 1399–1412. doi:10.1109/TVCG.2018.2867776. 2

[MBB*11] Marcus A., Bernstein M. S., Badar O., Karger D. R., Madden S., Miller R. C.: Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems* (Vancouver BC Canada, May 2011), ACM, pp. 227–236. doi:10.1145/1978942.1978975. 2

[MLR*11] Maciejewski R., Livengood P., Rudolph S., Collins T. F., Ebert D. S., Brigantic R. T., Corley C. D., Muller G. A., Sanders S. W.: A pandemic influenza modeling and vis. tool. *Journal of Visual Languages & Computing 22*, 4 (Aug. 2011), 268–278. doi:10.1016/j.jvlc.2011.04.002. 1, 2

[NYX*12] Nan Cao, Yu-Ru Lin, Xiaohua Sun, Lazer D., Shixia Liu, Huamin Qu: Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time. *IEEE Trans. on Vis. and Computer Graphics 18*, 12 (Dec. 2012), 2649–2658. doi:10.1109/TVCG.2012.291. 2

[SRMV16] Skianis K., Rossi M.-E. G., Malliaros F. D., Vazirgiannis M.: SpreadViz: Analytics and Vis. of Spreading Processes in Social Networks. In *2016 IEEE 16th Intl. Conf. on Data Mining Workshops (ICDMW)* (Barcelona, Spain, Dec. 2016), IEEE, pp. 1324–1327. doi:10.1109/ICDMW.2016.0195. 2

[ST20] Sulis E., Tambuscio M.: Simulation of misinformation spreading processes in social networks: an application with NetLogo. In *2020 IEEE 7th Intl. Conf. on Data Science and Advanced Analytics (DSAA)* (sydney, Australia, Oct. 2020), IEEE, pp. 614–618. doi:10.1109/DSAA49011.2020.00086. 2

[STP*17] Sun G., Tang T., Peng T.-Q., Liang R., Wu Y.: Socialwave: Visual analysis of spatio-temporal diffusion of information on social media. *ACM Trans. Intell. Syst. Technol. 9*, 2 (oct 2017). doi:10.1145/3106775. 2

[STTL15] Shi L., Tong H., Tang J., Lin C.: VEGAS: Visual influEnce GrAph Summarization on Citation Networks. *IEEE Trans. on Knowledge and Data Engineering 27*, 12 (Dec. 2015), 3417–3431. doi:10.1109/TKDE.2015.2453957. 2

[SWL*14] Sun G., Wu Y., Liu S., Peng T.-Q., Zhu J. J. H., Liang R.: EvoRiver: Visual Analysis of Topic Coopetition on Social Media. *IEEE Trans. on Vis. and Computer Graphics 20*, 12 (Dec. 2014), 1753–1762. doi:10.1109/TVCG.2014.2346919. 2

[VKPM15] Vallet J., Kirchner H., Pinaud B., Melançon G.: A Visual Analytics Approach to Compare Propagation Models in Social Networks. *Electronic Proceedings in Theoretical Computer Science 181* (Apr. 2015), 65–79. doi:10.4204/EPTCS.181.5. 2

[VLDBF15] Von Landesberger T., Diel S., Bremm S., Fellner D. W.: Visual analysis of contagion in networks. *Information Vis. 14*, 2 (Apr. 2015), 93–110. doi:10.1177/1473871613487087. 2

[WLY*14] Wu Y., Liu S., Yan K., Liu M., Wu F.: OpinionFlow: Visual Analysis of Opinion Diffusion on Social Media. *IEEE Trans. on Vis. and Computer Graphics 20*, 12 (Dec. 2014), 1763–1772. doi:10.1109/TVCG.2014.2346920. 2

[YDH*17] Yanez A., Duggan J., Hayes C., Jilani M., Connolly M.: PandemCap: Decision support tool for epidemic management. In *2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)* (Phoenix, AZ, Oct. 2017), IEEE, pp. 24–30. doi:10.1109/VAHC.2017.8387497. 2

[YWL*14] Yuan X., Wang Z., Liu Z., Guo C., Ai H., Ren D.: Vis. of social media flows with interactively identified key players. In *2014 IEEE Conf. on Visual Analytics Science and Technology (VAST)* (Paris, France, Oct. 2014), IEEE, pp. 291–292. doi:10.1109/VAST.2014.7042535. 1, 2

[ZCW*14] Zhao J., Cao N., Wen Z., Song Y., Lin Y.-R., Collins C.: #FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media. *IEEE Trans. on Vis. and Computer Graphics 20*, 12 (Dec. 2014), 1773–1782. doi:10.1109/TVCG.2014.2346922. 2