

Predicting the Evolution of Narratives in Social Media

Klaus Arthur Schmid¹, Andreas Züfle², Dieter Pfoser²,
Andrew Crooks², Arie Croitoru², and Anthony Stefanidis²

¹ Ludwig-Maximilians-Universität München
Institute of Informatics
`schmid@dbis.ifi.lmu.de`

² George Mason University
Department for Geography and Geoinformation Science
Center for Geospatial Intelligence
`azufler, dpfoser, acrooks2, acroitor, astefani@gmu.edu`

Abstract. The emergence of global networking capabilities (e.g. social media) has provided newfound mechanisms and avenues for information to be generated, disseminated, shaped, and consumed. The spread and evolution of online information represents a unique narrative ecosystem that is facilitated by cyberspace but operates at the nexus of three dimensions: the social network, the contextual, and the spatial. Current approaches to predict patterns of information spread across social media primarily focus on the social network dimension of the problem. The novel challenge formulated in this work is to blend the social, spatial, and contextual dimensions of online narratives in order to support high fidelity simulations that are contextually informed by past events, and support the multi-granular, reconfigural and dynamic prediction of the dissemination of a new narrative.

1 Introduction

Information technology of the 20th century allows members of social media communities to spread updates to friends and followers in real-time all over the world. Through this process information reaches a broader community while at the same time being re-shaped (i.e., altered, refined, or complemented). This represents a newfound example of communally curated narratives, reflecting the 21st century ethos of public sharing of information. We have observed this in numerous instances and at various scales, from the use of social media by the general public to provide timely information in the aftermath of the terrorist attacks in Boston and Paris, to the evolving narrative regarding election campaigns across the world, and the dissemination of information, hopes, and fears that follows the emergence of global epidemics like Zika. The spread and evolution of online information in such instances represent a unique narrative ecosystem that is facilitated by cyberspace but operates at the nexus of three dimensions:

- the social network dimension, as defined by the social networks that are formed between individual nodes and serve as potential dissemination routes,
- the spatial dimension, reflecting communication in the physical world, and
- the contextual dimension of the particular interests and opinions of these networks on diverse topics, providing context that discerns event responses.

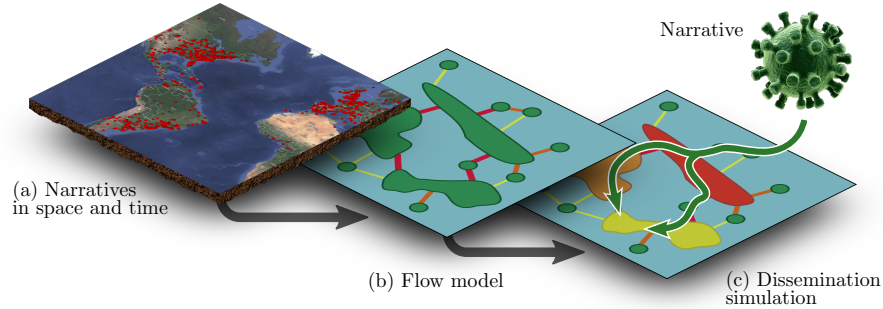


Fig. 1. Abstraction from Narrative occurrences to a flow model and simulation.

Current approaches to study this narrative ecosystem, e.g. [4, 6, 2, 1, 3], have primarily focused on the social network dimension of the problem. As a result, we have seen the development of complex network-based solutions to study, model, and compute information dissemination patterns as a function of the structure of the corresponding social networks. However, these approaches do not perform well when attempting to predict the dissemination of a narrative in the real-world. This is due to their lack of an understanding of the other dimensions of the communication process. The framework proposed in [8] considers the geo-spatial component, but without consideration of the underlying social network structure. To the best of our knowledge, no solutions exist that consider the contextual dimension to predict the dissemination of a narrative in space and time, thus learning how various narrative types tend to spread and evolve differently.

In order to develop more powerful models to predict the spread and evolution of narratives, we need to blend the social, spatial and contextual dimensions of online narratives in order to support high fidelity simulations that are contextually informed by past events, and support the multi-granular, reconfigural and dynamic nature of these networks. An overview of this vision is shown in Figure 1: First, Figure 1(a) illustrates the raw data, i.e., occurrences of a given narrative in space and time. For such narrative, a flow-model can be constructed by reducing the space of individuals relevant for a specific topic, as shown in Figure 1(b). Doing this for a large number of historic narratives yields a library of narrative dissemination models. This library can be used to predict the dissemination of a new narrative, by searching for similar historic narratives and using these to predict future dissemination.

2 State-of-the-Art

Traditionally, two types of models have been established to model diffusion and dissemination of information in networks: differential equation models [4] and agent based models [3]. An overview of these models can be found in [6]. A weakness of models using differential equations is the aggregation of individual agents into a relatively small number of compartments or populations. Within each population, people are assumed to be homogeneous and well mixed. Transitions among compartments are modeled as their expected value, losing important information about individual influencers and gate-keepers. In contrast, agent based models are able to capture heterogeneity between individuals, thus allowing to exploit the network structure, as well as individual attributes in the information diffusion model to improve diffusion prediction accuracy. Yet, such

models suffer from a high computational complexity, scaling up to hundreds of thousands of agents [3, 5]. Several models have been proposed to incorporate social and spatial information to detect current events [9, 11, 7], however, such work often lack the temporal component necessary to follow a narrative in time. Therefore, the ability to predict future information propagation patterns using past data remains a substantial scientific challenge.

3 Proposed Direction

In our work we aim to combine the high efficiency of differential equation models with the high modeling power of agent based models, by aggregating individuals to compartments/groups only as necessary. Towards such a hybrid model, we can reduce the problem complexity in two ways:

Reduction of the topic space: Different narratives share similar diffusion patterns in space as time, as we were able to show in preliminary work [8]. For instance, different topics related to entertainment disseminate in a similar way, whereas topics related to politics exhibit different dissemination patterns. Such clusters of similar narratives can be organized hierarchically, for example, a broader topic may be health, and under it we may have several subtopics, such as infectious diseases, or chronic diseases. These subgroups can be generated through the analysis of historical data (such as Twitter data). The resulting *library of narrative groups* represents abstractions of the information dissemination process, and can be fine-tuned in terms of its thematic resolution (moving from broader categories to more specific categories), in terms of its network resolution (moving from broader clusters of nodes to finer ones, as we will see in the next section). This allows us to balance computational performance and fidelity as desired for future simulations.

Reduction of the agent space: For a single specific topic, many users may have similar opinions, and can be aggregated into a population without significant loss of information. For instance, one user might be a vocal influencer and gate-keeper for the information dissemination of topics related to a specific type of sports, whereas this user may, at the same time, be oblivious to topics related to politics. Thus, for politics related topics, we may not need to model this user by an individual agent. This further observation allows to further reduce the space of users that need to be modelled. It allows to group of individuals that are oblivious to the given narrative on an aggregated level, while giving full detail to individuals that the model identifies to be trend-setters and vocal. We can model the information dissemination given (conditioned to) a specific topic this topic-archetype and apply an attributed graph clustering algorithm [10] to find communities of users which share a similar opinion towards the topic. This layer of abstraction, which is illustrated by the transition from Figure 1a) to Figure 1b), yields a set of abstracted information dissemination models for each narrative group.

Simulation and Prediction of Narrative Dissemination: Once we have a library of abstracted information dissemination models, these will be used for grounding an agent based simulation. In this simulation, an agent will be a person or a whole population, as dictated by the dissemination model. To start a

simulation, a new narrative is injected into the system. The first step of this grounding is to identify the narrative group of a new narrative. This identification task can be formulated as a supervised classification problem, mapping the new narrative to the most fitting narrative group in the narrative group library. Depending on the available information about the new narrative, this classification may be more or less detailed, thus allowing to dive more or less deep into the narrative group hierarchy, and thus, yielding a more or less detailed dissemination model. Intuitively, the longer we observe a new narrative, the more confident our model fitting will become.

4 Conclusion

The proposed framework would result in models that extend their power beyond the mere structure of the underlying social network. By learning the dissemination of past narratives, latent forms of spreading and evolving a narrative are also captured by this model: While we can not directly observe individuals sharing ideas physically, we can observe the consequence of both individuals frequently sharing the same ideas. This enables us to implicitly capture forms on information dissemination, and allow to learn features of individuals that are more likely, for the given narrative group, to pass on the ideas of others for further dissemination. In doing so we will have the potential to predict the dissemination of new narratives as they emerge.

References

1. E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *WWW*, pages 519–528. ACM, 2012.
2. M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *WWW*, pages 721–730. ACM, 2009.
3. I. N. Lympelopoulou and G. D. Ioannou. Online social contagion modeling through the dynamics of integrate-and-fire neurons. *Information Sciences*, 320:26–61, 2015.
4. V. Mahajan, E. Muller, and Y. Wind. *New-product diffusion models*, volume 11. Springer Science & Business Media, 2000.
5. B. Pires and A. T. Crooks. Modeling the emergence of riots: A geosimulation approach. *Computers, Environment and Urban Systems*, 61, Part A:66–80, 2017.
6. H. Rahmandad and J. Sterman. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science*, 54(5):998–1014, 2008.
7. T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pages 851–860. ACM, 2010.
8. K. A. Schmid, C. Frey, F. Peng, M. Weiler, A. Züfle, L. Chen, and M. Renz. Trendtracker: Modelling the motion of trends in space and time. In *SSTD@ICDM Workshop*, pages 1145–1152, 2016.
9. S. Unankard, X. Li, and M. A. Sharaf. Emerging event detection in social networks with location sensitivity. *World Wide Web*, 18(5):1393–1417, 2015.
10. Z. Xu, Y. Ke, Y. Wang, H. Cheng, and J. Cheng. A model-based approach to attributed graph clustering. In *SIGMOD*, pages 505–516. ACM, 2012.
11. X. Zhou and L. Chen. Event detection over twitter social media streams. *The VLDB journal*, 23(3):381–400, 2014.