Measuring Contagion on Social Media

SEBASTIAN SCHERR
KU Leuven, Belgium

Defining key terms related to social media contagion

Social networks can be defined as computer-mediated technologies that connect people virtually (online), creating digital networks that may exist for different purposes (e.g., personal or professional) and may have different affordances (e.g., ability to share content, make comments). Social media contagion is defined as a "social current" (from Durkheim's 1964 work on contagion), involving series of similar behaviors that are most likely reproduced by imitation or co-orientation, and that are typically observable at an aggregate level within digital social networks.

Contagion, of course, is a medical metaphor, as is the notion of "virality." Contagion implies that the mere contact with certain information is enough to infect people and immediately make them carriers and infectious spreaders of that information. Contagion has long been discussed and studied as a social phenomenon, even well before the emergence of social media. Researchers have noted that although contagion is a useful metaphor, the adoption of ideas and behaviors is tied to complex cognitive processes that tap into intentionality, trust, social norms, and human tendencies to confirm existing beliefs.

Despite this earlier work, contagion has undoubtedly become more of a focus of research in the context of social media, especially following the increase in digital research methods and the birth of computational social science (Lazer et al., 2009). Such work has tended to characterize the social media contagion phenomenon as (i) impermanent, (ii) developing over time, and (iii) evolving into specific social network dynamics. Social media contagion therefore reflects a human inclination to follow perceived social movements that link with individual predispositions and preferences.

Studies on social media contagion typically differentiate between affective (e.g., Ferrara & Yang, 2015) and behavioral contagions (e.g., Christakis & Fowler, 2007; Mønsted, Sapieżyński, Ferrara, & Lehmann, 2017), which are then sometimes further differentiated into simple contagion (i.e., observable contagion effects after one exposure) versus complex contagion (i.e., contagion effects only after multiple exposures).

Key questions and approaches with regard to social media contagion

Contagion dynamics on social media can be measured using the public pages of certain groups on social media through platform APIs (i.e., application programming interface that grants access to a platform's [user] data). Although data may be restricted by the individual privacy settings of platform users, public content can be accessed together with other platform activities (e.g., likes, comments, shares). Typically, researchers try to look for a set of coherent content from social media platforms and harvest it from the social media pages.

They may look at the lifetime of a post that can be stored as the temporal distance between the first and the last focal activity, or other indicators for the platform dynamics. This could be the number of users sharing the same content category, which provides an indicator for the stability of and the engagement with the content within user groups. Importantly, much of the research on the conditions that give rise to contagion focuses on ways to identify social media influencers. That is, contagion is particularly likely to occur for content that is posted (reposted, or recommended) by influencers on social media (people who have built up large online networks of people around them who regularly follow their posts). Because these individuals have such large networks of followers, content posted by them is much more likely to be shared rapidly with many others (and thus go viral).

Given this focus, various computational approaches have been developed to identify who, within a network, should be considered the social influencer(s) (Pei, Morone, & Makse, 2018). For example, a study could choose a random root user (e.g., the researcher him/herself) as a starting point to mine and generate his/her friends' list together with information about who they in turn are following (i.e., friends of friends, and so on) as well as their messages. A social network is thereby generated with users being nodes and their connections with followers (as well as people they follow) being the edges in the social network. Network interactions (e.g., referrals, retweets) are then transformed into weight information used to assess tie strengths. The influencer identification is achieved by complex calculations solving the problem of "optimal percolation"—that is, removing nodes from the network up to the point where the minimal set of nodes still keeps the connectivity of the whole network alive. The influencer nodes are those that bring the whole network toward imploding if eliminated. Of note, identifying several influencers that simultaneously exert influence on the social network (i.e., collective influence maximization) is a computationally challenging task of high practical relevance, for example to examine the spread (and success) of public health campaigns.

Above and beyond the impact of individual influencers, social media contagion has been described and studied in terms of a "rippling effect" (Christakis & Fowler, 2009). Research on the "Three Degrees of Influence Rule" (Christakis & Fowler, 2009, p. 28) suggests that our thoughts and behaviors line up with those of our friends (i.e., first degree), friends of our friends (i.e., second degree), and even their friends (i.e., third degree). The argument is that content shared online will ripple through our social network to the third degree, but not much further than this third degree.

Much of the research on contagion has focused on low-key habitual social media behaviors and low-cost decision-making processes such as scrolling through a social media timeline or choosing to read articles behind a link (e.g., Bond et al., 2012). However, online social media platforms can also influence offline attitudes and important behavioral decisions, thus spreading beyond the online platform, including mobility patterns of people in their leisure time (Leng, Dong, Moro, & Pentland, 2018).

Additionally, beyond simply depicting the social network structure, algorithm-based analyses are often undertaken to assess the positivity and negativity of sentiments within the shared content, or to identify communities within the network based on patterns of sharing information. For example, research on "homogeneous edges" of networks has examined how network homogeneity predicts the content's lifetime, including the emergence of so-called echo chambers in which clusters of like-minded individuals keep sharing their point of view with each other, thereby reinforcing their beliefs (Pariser, 2011). Researchers have also studied how far content characteristics are associated with the lifetime of a post. For example, Zollo and Quattrociocchi (2018) observed that fact-based, scientific content typically has a consistently high lifetime within small user groups, and more varied lifetime within larger groups. In contrast, they found that conspiracy-focused content (i.e., not based on true facts) tends to have a lifetime that increases steadily as the user group size increases as it is shared among like-minded individuals and ripples through the network.

Untangling contagion from homophily

It is important to note that individuals' characteristics and their behaviors are often correlated with network structure due to homophily. That is, over time, people surround themselves with similar others (i.e., assortative mixing), and thus, similar attitudes and behaviors may be an expression of social contagion, but most likely (more than 50%) they are simply expressing homophily and not each other's influence (see Aral, Muchnik, & Sundararajan, 2009). Therefore, it is important that homophily effects are controlled for when assessing contagion, to avoid upward estimation bias (see Aral et al., 2009).

To help deal with this issue, researchers often use matching techniques to compare individuals with a different treatment status (e.g., nodes in a network that have adopted a technology vs. nodes that did not adopt it), but with similar individual characteristics, and similar network qualities (e.g., the number of "adopter friends" in their social network). Over time matched sample tests are performed (e.g., matched on a daily, weekly, biweekly, etc. basis) comparing treated versus nontreated (but otherwise similar) individuals with a similar social network. The idea is that differences can then be explained by individual characteristics or network qualities, thereby speaking to contagion effects (see Aral et al., 2009).

Specific methodological and ethical considerations

What are some key methodological and ethical considerations of this type of research? Despite the new and exciting possibilities of researching contagion in digital environments, the hosting platforms are private, for-profit companies (e.g., Twitter, Sina Weibo, Facebook, Instagram, WeChat). As such, they can decide to grant data access only to a limited group of researchers or demand coauthored collaborations in exchange for researchers having access to data or change the availability or structure of their user data at any time. Moreover, these platforms continuously change their appearance to maximize their users' time spent on the platform. All of this not only challenges the individual work of researchers who have designed their work (sometimes over years!) to particular APIs and data structures, but it also presents important challenges for the replicability of findings. Somewhat relatedly, there are real issues about the generalizability of many data sets, given that users with certain privacy settings might not be included in the data.

Additionally, there are various important questions about whether ethical standards for research on human subjects are being applied with online data, and whether data-privacy regulations are sufficient to meet research ethics needs. One approach has been to request completely de-identified data that cannot be traced back to any individual (e.g., for network analysis). Another has been to argue that the platform's terms of service sort of provide the participants' informed consent for participation in research (see Kahn, Vayena, & Mastroianni, 2014). Indeed, platforms routinely manipulate what some viewers see (called A/B testing) to tweak their format. However, it remains unclear whether such user agreements truly constitute informed consent to participate in large-scale social experiments in which academics and private platforms would manipulate the content shown to individuals without their knowing ahead of time or being told afterwards. In one of the more controversial examples, Facebook and researchers manipulated the affective tone of people's social media feed to observe online effects (see Kramer, Guillory, & Hancock, 2014). Ethical concerns become particularly pressing if the research hypothesis assumes treatment effects (e.g., a higher voter turnout in a group that sees particular content online; see Jones, Bond, Bakshy, Eckles, & Fowler, 2017), leaving the control or comparison group without access to the beneficial content. Some have argued that, given the ethical complexity of this ever-increasing body of research, journals serve crucial roles as gatekeepers, ethical supervisors, and research quality assessors (Kahn et al., 2014).

Finally, an alternative to collaborating with companies or accessing platform data through their (controlled) APIs is to engage in automated web scraping. Researchers can harvest data from social media automatically and in line with the terms of service, before further analyzing them. One recent example comes from nonsuicidal self-injury (NSSI) on social media. Scherr, Arendt, Frissen, and Oramas (2019) scraped pictures posted with self-harm-related hashtags from Instagram in line with the terms of service of the platform and then automatically analyzed the prevalence of these behaviors in the pictures, using an automatic image-recognition algorithm. Automatic solutions could also address the ethically problematic situation when human coders manually code highly problematic content.

SEE ALSO: Big Data, Collection of (Social Media, Harvesting); Measurement of Media Exposure; Measuring Behavior in Media Psychology; Unobtrusive Measures for Media Research

References

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America, 106*(51), 21544–21549. doi:10.1073/pnas.0908800106

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature, 489*(7415), 295–298. doi:10.1038/nature11421 PMID: 22972300

Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine, 357*(4), 370–379. doi:10.1056/NEJMsa066082

Christakis, N., & Fowler, J. (2009). *Connected: The surprising power of our social networks and how they shape our lives*. New York, NY: Little Brown and Company.

Durkheim, E. (1964) *The rules of sociological method*. New York, NY: The Free Press.

Ferrara, E, & Yang, Z (2015). Measuring emotional contagion in social media. *PLoS One, 10*(11): e0142390. doi:10.1371/journal.pone.0142390

Jones, J. J., Bond, R. M., Bakshy, E., Eckles, D., & Fowler, J. H. (2017). Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 U.S. presidential election. *PLoS One, 12*(4), e0173851. doi:10.1371/journal.pone.0173851

Kahn, J. P., Vayena, E., & Mastroianni, A. C. (2014). Opinion. Learning as we go: Lessons from the publication of Facebook's social-computing research. *Proceedings of the National Academy of Sciences of the United States of America, 111*(38), 13677–13679. doi:10.1073/pnas.1416405111

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788. doi:10.1073/pnas.1320040111

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... Van Alstyne, M. (2009). Life in the network: The coming age of computational social science. *Science, 323*(5915), 721.

Leng, Y., Dong, X., Moro, E., & Pentland, A. S. (2018). The rippling effect of social influence via phone communication network. In S. Lehmann & Y.-Y. Ahn (Eds.), *Complex spreading phenomena in social systems: Influence and contagion in real-world social networks* (pp. 323–333). Cham, Switzerland: Springer.

Mønsted, B., Sapieżyński, P., Ferrara, E., & Lehmann, S. (2017). Evidence of complex contagion of information in social media: An experiment using Twitter bots. *PLoS One, 12*(9): e0184148. doi:10.1371/journal.pone.0184148

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London: Penguin.

Pei, S., Morone, F., & Makse, H. A. (2018). Theories for influencer identification in complex networks. In S. Lehmann & Y.-Y. Ahn (Eds.), *Complex spreading phenomena in social systems: Influence and contagion in real-world social networks* (pp. 125–148). Cham: Springer.

Scherr, S., Arendt, F., Frissen, T., & Oramas, J. (2019). Detecting intentional self-harm on Instagram: Development, testing, and validation of an automatic image-recognition algorithm to discover cutting-related posts. *Social Science Computer Review*. Advance online publication. doi:10.1177/0894439319836389

Zollo, F., & Quattrociocchi, W. (2018). Misinformation spreading on Facebook. In S. Lehmann & Y.-Y. Ahn (Eds.), *Complex spreading phenomena in social systems: Influence and contagion in real-world social networks* (pp. 177–196). Cham: Springer.

Further reading

Goldberg, A., & Stein, S. K. (2018). Beyond social contagion: Associative diffusion and the emergence of cultural variation. *American Sociological Review, 83*(5), 897–932. doi:10.1177/0003122418797576

Sebastian Scherr (PhD, University of Munich) is a tenure-track Assistant Professor at the School for Mass Communication Research, KU Leuven, Belgium. His research interests focus on individual and structural susceptibility factors for media effects in the domains of health and political communication, with special emphasis on mental health and suicide prevention through media, as well as quantitative, qualitative, and computational methods.