

# Collective behaviors and networks

Giovanni Luca Ciampaglia, Emilio Ferrara\* and Alessandro Flammini

\*Correspondence:  
ferrarae@indiana.edu  
School of Informatics and  
Computing, Indiana University,  
Bloomington, USA

## Abstract

The goal of this thematic series is to provide a discussion venue about recent advances in the study of networks and their applications to the study of collective behavior in socio-technical systems. The series includes contributions exploring the intersection between data-driven studies of complex networks and agent-based models of collective social behavior. Particular attention is devoted to topics aimed at understanding social behavior through the lens of data about technology-mediated communication. These include: modeling social dynamics of attention and collaboration, characterizing online group formation and evolution, and studying the emergence of roles and interaction patterns in social media environments.

## 1 Introduction

The rise in usage of online social media provides a wealth of information about social phenomena and human behavior at scale, at least to the extent to which interactions, intentions and beliefs measured online reflect their real-world counterparts [1, 2]. Data about online traces of activity from Twitter, Facebook, Wikipedia, blogs, etc., have been used to predict elections and political opinions [3, 4], movie revenues at the box office [5, 6], fluctuations in the stock market [7], and the spreading of influenza [8–12], to cite a few examples. Similar data have shed light on the mechanisms behind social influence [12, 13] and spread of behavior [14, 15], or to study the diffusion of viral information [16–18], and the dynamics of social protests [19, 20].

Even preceding the surge of scholarly attention toward social media and online social networks, the use of Agent-based Models (ABM) has grown in scope, percolating to several disciplines within the social sciences from economics to environmental policy, sociology, and psychology. One major appeal of the agent based approach is the possibility to test *in silico* hypotheses about the emergence of macroscopic behavior as result of simple interaction rules among stylized agents [21, 22]. In recent years the focus has started to shift from testing the plausibility of specific theories to the development of quantitatively accurate models; rather than just testing ?what-if? scenarios, ABMs are now being used to provide quantitative forecasts in social systems.

Although the Network Science community, which studies interconnected socio-technical systems, and the ABM community, which simulates artificial societies as groups of interacting agents, have similar focus and a large overlap in interests, they are still separated from a profound chasm in their methodological approach. We believe that each community would greatly benefit by a larger degree of acquaintance with the methodological approaches of the other. Agent-based models that are strongly informed by empirical

facts and capable of producing predictions at multiple scales and resolutions - predictions for which empirical data are presently available - could be a highly desirable outcome of the increased interaction among the two communities.

These are the driving motivations behind our series of workshops *Covenant: Collective Behaviors and Networks?* that have been held in conjunction with the European Conference and Complex Systems (ECCS). The great success of the first edition, which was held in Barcelona (Spain) in September 2013, set a milestone by bringing together these two communities, and brought us to expand even further our objectives for a second edition, held in Lucca (Italy) in September 2014. This event replicated the success of the first and surpassed our wildest expectations, with a peak attendance of a hundred participants, and close to 50 original submissions by researchers and practitioners from all over the world.

The goal of the present thematic series is twofold: showcase the most outstanding contributions presented at these two meetings, and provide a discussion venue about recent advances in the study of networks and their application to the study of collective behaviors. The first five contributions published here have been carefully selected among those presented at Covenant 2013, and they present advances in three areas: (i) modeling social dynamics of attention [23] and collaboration [24]; (ii) characterizing online group formation and evolution [25]; (iii) studying the emergence of sharing habits patterns [26] and roles [27] in social media environments.

## 2 Contributions

The first contribution, by Ruiz *et al.* [23], investigates the dynamics of content production in an online microblogging community, and in particular the interplay between user activity and the attention she receives. In online social network (OSN) media content (such as photos, stories, news, etc.) is produced by the same set of people. As a consequence, the evolution of OSN sites is driven by the complex interplay between individual activity and attention received from others. This has important implications for the online communities. Receiving attention is a non-monetary reward that is crucial to sustain user engagement and prevent churn; therefore understanding what strategies are employed by the most successful users is likely of interest to anybody who wishes to promote socially sustainable communities, both online and offline. Proxies for collective attention are easy to measure in the digital world, and several works have approached the issue from different angles. Here the authors analyse a novel and interesting dimension of collective attention, the efficiency, defined as the ratio between the volume of collective attention received and the volume of content produced by a single user. They find that 56% of users in the system have very well-defined efficiency patterns over time, exhibiting either an increasing/decreasing, or peaking behavior. Further analyses lead the authors to conclude that increases in efficiency are determined by the creation of high-quality content, but that the attention acquired in this way has to be sustained by means of social exchanges (such as commenting or liking) to maintain high efficiency. Whenever this form of social activity is missing, efficiency quickly drops.

The second paper, contributed by Iñiguez *et al.* [24], also looks at a content-producing online community. In contrast to Ruiz *et al.* [23] that focused on a system, where contents is exchanged over social connections, here the authors investigate the free online encyclopedia Wikipedia, a strictly collaborative environment, where social connections arguably play a lesser role. Wikipedia is famous for allowing everyone to alter its content. This policy of low participation barrier has resulted in a surprisingly fast growth in its first decade

of existence. The other side of the coin is, however, the occurrence of conflicts among contributors with differing viewpoints. Wikipedia, with its detailed records about the history of edits on each article, provides an ideal ground to study how conflict arise and are solved. Indeed data about conflicts in other systems are usually difficult to find, possibly for the negative social connotation of the subject. Iñiguez *et al.* make an original contribution to this line of research by proposing a stylized agent-based model of fictitious Wikipedia editors that compete for control of an article (the medium). All editors and the medium are endowed with an internal opinion-like variable. For editors this represents their own viewpoint on the topic of the medium, and for the medium this can be thought as the most recent viewpoint contributed to it. Opinions are continuous variables and their dynamics follows the so-called Bounded Confidence (BC) rule from opinion dynamics. Such a stylized model, while simplifying several important aspects of Wikipedia's editorial process, still features, as the authors report, a rich dynamic. In particular, different regimes, corresponding to empirical observation of conflict on real Wikipedia pages, can be found for different ranges of key model parameters.

The third work, by Martin-Borregon *et al.* [25], studies Flickr, the popular photo-sharing platform, to understand how social groups form and evolve in time, space, and across the socio-topical dimension. The authors propose a general model to characterize groups through several metrics of reciprocity, activity and topical diversity (which embody the theory of *common identity and common bond*). The model clusters groups according to their temporal activity into three categories: evergreen, short-lived, and bursty ones. The authors' analysis shows that their model predicts accurately the type of a groups when compared with the manually-generated ground truth. The model also demonstrates that: (i) geographically-wide groups are longer-lived than local ones; (ii) topical groups are more robust to user churn than other types and tend to exhibit constant activity; and, (iii) social groups have bursty activity patterns, with most members joining at the beginning and then interacting only occasionally. The definition of groups according to this framework provides a more nuanced description of community if compared with that obtained solely by clustering the user social graph, better capturing user behaviors and group activities. In fact, the authors show that groups identified by the framework and clusters obtained from community detection don't overlap much, and are more often social than the declared ones. This agrees with the increasing body of literature that highlights the limitations of traditional topology-based network clustering to identify dynamical characteristics of socio-technical systems. The work finally concludes that information diffusion is affected by the grouping, with social and bursty groups spreading information across the boundaries more efficiently than topical and evergreen ones.

The fourth and fifth contributions selected from "Covenant 2013" both make use of Twitter data to study socio-technical environments: the former aims to model content sharing habits, and the latter at understanding the emergence of roles on the platform.

The work by An *et al.* [26] explores four different dynamics that contribute to the sharing behavior of news on social media: gratification, selective exposure, socialization, and trust. Traditional literature explored these dimensions independently, and without making use of datasets containing real social interactions and behaviors at scale. An *et al.* explore in particular whether the theories of *selective exposure* and *echo chamber* can be observed in a non-controlled environment (as done in the past by traditional psychologists). They also discuss what factors drive users to predominantly consume information that is aligned

with their pre-existing views. The work focuses on political news, as political leaning has been identified as one of the critical factors making people with different viewpoints drift apart. The authors first collected a large longitudinal dataset of tweets - produced during eight months in 2009 - from users sharing links to news articles selected from a list of 22 news agencies. They propose a model based on 12 features called PoNS (*Political News Sharing*), featuring the four social factors mentioned above. They then carefully design a protocol to evaluate their model. In particular, they determine which features contribute the most to the likelihood of rebroadcasting (retweeting) a news item based on its political inclination (either in line or opposing one's political views), and provided that the news story comes from an *official channel* or through friends. After adjusting for various confounding factors and possible sources of bias, they observe that homophily strongly limits who connects to whom on Twitter. Users are disproportionately more likely to connect with and retweet from others who share their same political views. The emergence of this polarization effect, which was first observed for Twitter in previous studies [28], makes it difficult for ideas to spread from one group to another. This work also highlights for the first time that individuals are much more likely to retweet news that oppose their views when they come from their contacts, compared to official accounts. This contrasts the broadly regarded theory of *cognitive dissonance*, which posits that individuals tend to stick even more to their views when they are faced with opposing ones. Finally, the authors ranked the predictive power of the twelve features they selected. They find that, when one controls for the number of exposures, other variables become important, although not uniformly for all users: some prefer popular stories, while others value those coming from trusted friends.

The fifth and last contribution, by González-Bailón *et al.* [27], focuses on the sociological theory of *brokerage* and extends it to complex, large-scale social networks. The goal of this work is to provide a method to identify user roles in inter-personal communication dynamics, leveraging information about network topology and, in particular the community structure that characterizes online social networks like Facebook and Twitter [29, 30]. The assumption is that roles respond to a division of labor that reflects different functions or behaviors, within the network. Network features, the authors hypothesize, may help detecting structurally similar positions and communication dynamics. The authors collect data on discussions about the Spanish political protest of May 2012 from Twitter, and pair this data with the classic Zachary's Karate Club dataset as a baseline. The question they aim to answer is whether individuals who exhibit similar network roles also behave similarly in terms of communication patterns. Additionally, they identify the most significant roles in this context. The authors proposed a hybrid local-global method (HM) based on clustering the network via a community detection algorithm. Once the community structure is obtained, the two classic schemes of GF (local role inference based on paths of length two around each node) and GA (global role inference based on the density of the membership community) can be computed and combined for each node. Based on their analysis, the authors observe that in Twitter most individuals play a representative role, are peripheral and exhibit low levels of brokerage, both at the local and global level. There is little opportunity for information to flow from community to community, and that most users don't control direct diffusion channels. The work finally shows that similar roles trigger similar behavior. By measuring number of retweets and mentions received by each user as proxy for authority and salience, the authors show that local and

global brokers are consistently the most retweeted users (therefore, essential conduits for global information flow), while the most salient users are brokers only at local level.

### 3 Conclusions

Technology-mediated social collectives are taking an important role in the design of social structures. Yet, our understanding of the complex mechanisms governing networks and collective behavior is still shallow. Social systems are often viewed as instances of entirely unpredictable systems, their future bound to be dominated by contingency and happenstance. Agent-based models and complex networks methods are relatively novel items in the toolbox of computational social scientists, and as any tool they have their strengths and limitations. In this thematic series we aimed to show that combining these two approaches may help formulate better models and perhaps open up the possibility of entirely novel modeling approaches. Thank to the ubiquity of big data about social interactions, computational models of social phenomena are becoming more accurate in forecasting complex social phenomena. We foresee an intense growth of synergistic interactions between ABM and network based approaches in the future of computational social science.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Received: 15 December 2014 Accepted: 15 December 2014 Published online: 30 December 2014

#### References

- Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D, Van Alstyne M, et al (2009) Life in the network: the coming age of computational social science. *Science* 323(5915):721
- Vespignani A (2009) Predicting the behavior of techno-social systems. *Science* 325(5939):425
- Metaxas PT, Mustafaraj E (2012) Social media and the elections. *Science* 338(6106):472-473
- Conover MD, Gonçalves B, Flammini A, Menczer F (2012) Partisan asymmetries in online political activity. *EPJ Data Sci* 1:6
- Asur S, Huberman BA (2010) Predicting the future with social media. In: 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (WI-IAT), Vol 1. IEEE, pp 492-499
- Mestyán M, Yasseri T, Kertész J (2013) Early prediction of movie box office success based on Wikipedia activity big data. *PLoS ONE* 8(8):e71226
- Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. *J Comput Sci* 2(1):1-8
- Culotta A (2010) Towards detecting influenza epidemics by analyzing Twitter messages. In: Proceedings of the first workshop on social media analytics. ACM, New York, pp 115-122
- Signorini A, Segre AM, Polgreen PM (2011) The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS ONE* 6(5):e19467
- Chew C, Eysenbach G (2010) Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS ONE* 5(11):e14118
- Lazer D, Kennedy R, King G, Vespignani A (2014) Big data. The parable of google flu: traps in big data analysis. *Science* 343(6176):1203
- Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092):337-341
- Bond RM, Fariss CJ, Jones JJ, Kramer AD, Marlow C, Settle JE, Fowler JH (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295-298
- Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194-1197
- Golder SA, Macy MW (2011) Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333(6051):1878-1881
- Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854-856
- Szabo G, Huberman BA (2010) Predicting the popularity of online content. *Commun ACM* 53(8):80-88
- Weng L, Menczer F, Ahn YY (2013) Virality prediction and community structure in social networks. *Sci Rep* 3:2522
- Conover MD, Davis C, Ferrara E, McKelvey K, Menczer F, Flammini A (2013) The geospatial characteristics of a social movement communication network. *PLoS ONE* 8(3):e55957
- Conover MD, Ferrara E, Menczer F, Flammini A (2013) The digital evolution of Occupy Wall Street. *PLoS ONE* 8(5):e64679
- Vespignani A (2012) Modelling dynamical processes in complex socio-technical systems. *Nat Phys* 8(1):32-39
- Epstein JM (1996) Growing artificial societies: social science from the bottom up. Brookings Institution Press

23. Ruiz CV, Aiello LM, Jaimes A (2014) Modeling dynamics of attention in social media with user efficiency. *EPJ Data Sci* 3:5
24. Iñiguez G, Török J, Yasseri T, Kaski K, Kertész J (2014) Modeling social dynamics in a collaborative environment. *EPJ Data Sci* 3:7
25. Martin-Borregon D, Aiello LM, Grabowicz P, Jaimes A, Baeza-Yates R (2014) Characterization of online groups along space, time, and social dimensions. *EPJ Data Sci* 3:8
26. An J, Quercia D, Cha M, Gummadi K, Crowcroft J (2014) Sharing political news: the balancing act of intimacy and socialization in selective exposure. *EPJ Data Sci* 3:12
27. González-Bailón S, Wang N, Borge-Holthoefer J (2014) The emergence of roles in large-scale networks of communication. *EPJ Data Sci* 3:32
28. Conover M, Ratkiewicz J, Francisco M, Gonçalves B, Menczer F, Flammini A (2011). Political polarization on twitter. In: ICWSM
29. Ferrara E (2012) A large-scale community structure analysis in Facebook. *EPJ Data Sci* 1:9
30. De Meo P, Ferrara E, Fiumara G, Provetti A (2014) On Facebook, most ties are weak. *Commun ACM* 57(11):78-84

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---