Let's talk about refugees: Network effects drive contributor attention to Wikipedia articles about migration-related topics

Jürgen Lerner[1] and Alessandro Lomi[2]

[1] University of Konstanz, Germany, juergen.lerner@uni-konstanz.de
[2] University of Lugano, Switzerland, alessandro.lomi@usi.ch

Abstract. Contributions by voluntary users are one of the most crucial resources in the online encyclopedia Wikipedia. In this paper we propose relational event models to analyze dynamic network effects explaining the allocation of contributor attention to Wikipedia articles about migration-related topics. Among others, we test for the presence of a rich-get-richer effect in which articles edited by many users are likely to receive even more contributions in the future and uncover which users start working on less popular articles. We further analyze local clustering effects in which pairs of users tend to repeatedly collaborate on the same articles as well as interaction between contributions to encyclopedic articles and engagement in associated talk pages. We demonstrate that these network effects that regulate collaborative work in Wikipedia act over and above general popularity of the articles' topics as revealed by the number of pageviews.

Keywords: social network analysis, relational event model, online peer-production, attention, Wikipedia

1 Introduction

Attention is one of the most valuable resources in our contemporary information-rich societies [33, 34, 29]. This is particularly the case for online peer-productions, such as open source software, or the user-generated encyclopedia Wikipedia, in which projects compete for attention of volunteer contributors [22, 10, 11, 2, 21]. Due to the lack of predefined central coordination mechanisms and the absence of explicit monetary incentives, we argue that contributor attention is driven, at least in part, by emergent collaboration networks arising from task-oriented interaction among participants [14, 36, 20].

From an analytical standpoint, online peer-productions have the desirable property of providing complete and fine-grained data availability [18]. For instance, in Wikipedia – the empirical setting of this paper – we know the exact point in time in which any user contributes to any article or engages in any related project page. Yet, current analytical approaches are unable to benefit from the information afforded by this high level of resolution. The typical approach

involves forms of aggregation of editing events either over time intervals and/or over users or articles [35, 14, 24].

Using the new `eventnet`[3] software, in this paper we propose and implement an analysis of collaborative work in Wikipedia via relational event models (REM) [5]. REM can specify and estimate time-varying contribution rates separately for each user-article pair and can thereby uncover network effects explaining *who contributes when to which article*, keeping the full granularity of Wikipedia log data, that is, aggregating neither over time intervals nor over users or articles. Among others we attempt to shed light on the following research questions.

- Do we observe a "rich-get-richer" effect [1] such that articles that received already many contributions in the past are also more attractive in the future?
- If so, who (if anyone) starts working on the less popular articles?
- Do we find evidence for local clustering, such that pairs of users who collaborated on the same articles in the past are also more likely to collaborate on potentially different articles in the future?
- Is there systematic interdependence between conributions to articles and engagement in related talk pages? Does discussion tend to precede article writting [31] or is it rather the other way round?
- Are such network effects (if any) just a reflection of varying general popularity of the articles' topics as measured by the number of pageviews?

We test these and other hypothetical effects on the network of relational events encoding user contributions to Wikipedia articles on migration-related topics (see Sect. 2.1 for the definition of this set of articles). We consider this setting as highly relevant, since migration is a topic seen by citizens of many countries as one of the top issues facing politics and society.[4] Migration is also a highly representative example of a topic that is attracting considerable public attention. As such, migration-related topics are particularly appropriate to study the interplay between exogenous popularity (caused by the attention provided by the general public) and endogenous popularity (caused by the attention provided by contributing users of Wikipedia).

1.1 Background and further related work

Receiving attention by contributing users is one of the most crucial resources in online peer-production. For instance, it is a strong predictor for the quality of Wikipedia articles [21]. In the reverse direction, [27] have shown that attention in a cultural marketplace is rather unpredictable and that users' knowledge about other users' preferences has more influence than the quality of products.

Distributions of user activity and attention to articles in Wikipedia over time and (geographic) space [13, 35, 8] have been mostly analyzed by aggregating over time, users, or articles. A sequence analysis of editing patterns has

[3] http://algo.uni-konstanz.de/software/eventnet/
[4] Eurobarometer 89. http://ec.europa.eu/commfrontoffice/publicopinion/

been proposed by [16] and global characteristics of the "who-edits-after-whom" network have been analyzed in [15]. In contrast, in our work we analyze the two-mode network connecting users to the articles they contribute to. This two-mode network has been modeled with exponential random graph models (ERGM) by [14]. Yet, ERGMs require to aggregate dyadic interaction over time, loosing the fine-grained time information of Wikipedia log data.

Relational event models (REM) [5, 4] are specifically designed for networks of interaction events that are observed in (near-)continuous time, such as computer-mediated communication networks – or contributions of Wikipedia users to articles. REM specify separatly for each dyad (i. e., pair of nodes) a time-varying event rate and can thereby exploit the full time-granularity of data stemming from computer-mediated interaction [7, 25, 19, 28]. Yet, to our knowledge, REM have not been used before to model the allocation of user attention to Wikipedia articles, as it is done in our paper.

Engagement of Wikipedia users in talk pages (that is, pages providing room for discussing issues about article writting [31]) has also been analyzed in previous work. Activity in talk pages has been related to article quality [26] and compared over different languages [23]. Further work analyzed the dynamics of talk [12] and proposed a model predicting which posts are likely to receive replies [9]. In our paper we propose relational event models for user engagement in editing and discussion, where editing can potentially depend of previous editing and talk and vice versa.

2 Relational event models for the Wikipedia network

In this section we first describe the input data as a sequence of time-stamped and typed dyadic events and define the *network of past events* which is a dynamic data-structure, encoding past interaction on pairs of nodes. We then recall a general framework to specify and estimate models for dyadic event rates, than is, for the intensity of contributions of particular users to particular articles and propose specific time-varying explanatory variables (dyadic *statistics*) used to specify concrete models in our paper, see Figs. 1 and 2 for illustration.

2.1 Data

Our sample of articles consists of all articles from the English-language edition of Wikipedia that are in the category `Human migration`[5] or in a sub-category of it or in a sub-category of a sub-category. Our sample of *contributing users* consist of all registered users editing any of these articles, or making any edit to an associated talk page. Our sample of *dyadic events* is a list $E = (e_1, \ldots, e_N)$, where each event has the form

$$e_i = (u_i, a_i, t_i, x_i) \ ,$$

[5] https://en.wikipedia.org/wiki/Category:Human_migration

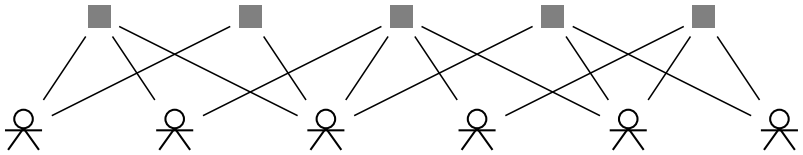

Fig. 1. Bipartite event network connecting Wikipedia users (bottom) to the articles (top) they contribute to. For a user u, an article a, and a time point t, we define two edge weights on (u, a) at t as the number of events of type *edit*, respectively *talk*, that u performed on a at any point in time $t' < t$ strictly before t. The rates of events of both types on (u, a) at t are specified as functions of these weights on (u, a) or on surrounding edges, as well as by exogenous properties of the article a (see text and Fig. 2).

encoding that user u_i performed an event of type x_i on article a_i at time t_i. The event type x_i is either *edit* (if u_i edited the article a_i at t_i) or *talk* (if u_i edited the talk page associated with article a_i at t_i). Time is given by the second and the observation period is from January 15, 2001 (the launch of Wikipedia) to January 1st, 2018 (the time of data collection). Events are sorted in non-decreasing order with respect to time, that is, it holds $t_1 \leq t_2 \leq \ldots \leq t_N$. This event network consists of more than 4,000 articles and 87,000 users connected by more than 950,000 dyadic events.

Besides information about user contributions to articles we use the number of *pageviews*[6] (by any Internet user, whether registered in Wikipedia or not) to articles in our sample. The number of pageviews are given separately for every article and every hour and are interpreted in this paper as a dynamically changing measure of general interest in the article's topic. Since Wikimedia's definition of pageviews changed in May 2015, we fit models that consider pageviews on the reduced observation period from May 1st, 2015 to January 1st, 2018.

2.2 The network of past events

Drawing on ideas from [4] we define the *event network* $G[E]$ associated with the event sequence E to be a dynamic, weighted two-mode network. For a time point t, the two node sets are the set of users U_t comprising all users who have initiated at least one event at or before t and the set of articles A_t comprising all articles who have received at least one event at or before t.

Two dynamically changing weight functions defined on user-article pairs encode the number of past events of type *edit*, or *talk*. In formulas, it is for a time point t, a user $u \in U_t$, and an article $a \in A_t$

$$past.edit(u, a, t) = |\{(u', a', t', x') \in E\colon\ t' < t \wedge u' = u \wedge a' = a \wedge x' = edit\}|$$
$$past.talk(u, a, t) = |\{(u', a', t', x') \in E\colon\ t' < t \wedge u' = u \wedge a' = a \wedge x' = talk\}|$$

[6] https://dumps.wikimedia.org/other/analytics/

For a time point t we use the symbol $G[E;t]$ to denote the *network of past events* which is a two-mode network with node sets U_t and A_t and two edge weights $past.edit(\cdot,\cdot,t)$ and $past.talk(\cdot,\cdot,t)$, both defined on $U_t \times A_t$. The edge set $E_t \subseteq U_t \times A_t$ is implicitly defined to consist of all pairs (u,a) for which $past.edit(u,v,t) > 0$ or $past.talk(u,v,t) > 0$. (By a slight abuse of notation we use the symbol E for dyadic events and for edges. This should not cause any confusion.) We emphasize that the two weight functions of $G[E;t]$ are functions of events that happen strictly before t (not of events that happen at t).

2.3 A framework for modeling dyadic, typed events

Our model for dyadic, typed events fits into the framework proposed by [5]. Let t be any time point, $(u,a) \in U_t \times A_t$ be any user-article pair, and let $x \in \{edit, talk\}$ be one of the two event types. Denote by $T \geq t$ the random variable for the time of the next event of type x on the dyad (u,a), given the network of past events $G[E;t]$. The time-varying *hazard rate* for events of type x on (u,a) at t is defined as

$$\lambda(u,a,x,t;G[E;t]) = \lim_{\Delta t \to 0} \frac{Prob(t \leq T < t + \Delta t \,|\, t \leq T; G[E;t])}{\Delta t} \ .$$

The hazard rate λ can be interpreted as the expected number of events in a time interval of length one [17]. Thus it is also called event intensity or frequency.

Adopting the Cox proportional hazard model [6, 17] (which corresponds to the ordinal model in [5]) we specify the hazard rate by the general functional form

$$\lambda(u,a,x,t;G[E;t];\theta^{(x)}) = \lambda_0(x,t) \cdot \exp\left(\sum_{j=1}^{k} \theta_j^{(x)} \cdot s_j(u,a;G[E;t])\right) \ , \qquad (1)$$

where $\lambda_0(x,t)$ is a time-varying baseline event rate for all dyads in the network, the $s_j(u,a;G[E;t])$, for $j = 1,\ldots,k$, are explanatory variables (*statistics*) that are functions of the network of past events $G[E;t]$ around user u and article a (see Sect. 2.5 and Fig. 2), and $\theta^{(x)} = (\theta_1^{(x)},\ldots,\theta_k^{(x)})$ are real-valued parameters (to be estimated; see below) revealing which statistics tend to increase or decrease the event rate.

Let $e = (u,a,t,x) \in E$ be any observed event. In the Cox proportional hazard model [6] (also in the ordinal model in [5]), the probability that the event of type x at time t happens on (u,a), rather than on any other dyad in the network, is

$$Prob^{(x)}(e;G[E;t];\theta^{(x)}) = \frac{\exp\left(\sum_{j=1}^{k} \theta_j^{(x)} \cdot s_j(u,a;G[E;t])\right)}{\sum_{(u',a') \in U_t \times A_t} \exp\left(\sum_{j=1}^{k} \theta_j^{(x)} \cdot s_j(u',a';G[E;t])\right)} \qquad (2)$$

Assuming that events are conditionally independent, given the network of past events, we obtain two joint probability functions for events of type $x \in \{edit, talk\}$ in the given event sequence $E = (e_1, \ldots, e_N)$

$$Prob^{(x)}(E; \theta^{(x)}) = \prod_{e_i \in E:\, x_i = x} Prob^{(x)}(e_i; G[E; t_i]; \theta^{(x)}) \ .$$

Parameters $\theta^{(x)}$ are estimated to maximize the likelihood $L(\theta^{(x)}) = Prob^{(x)}(E; \theta^{(x)})$. However, trying to do so would cause severe runtime problems, as described in Sect. 2.4.

2.4 Parameter estimation under sampling

The most computationally intensive part in computing (or maximizing) the likelihood is the denominator in Eq. (2), which is a sum over all user-article pairs on which the event at time t could have happened. In our data, we have at the end of the observation period more than 350 million such pairs. Since Eq. (2) has to be computed for nearly 1 million events, this implies an unfeasible runtime.

A solution is provided by case-control sampling [3] which is often applied in epidemiological studies of rare diseases. Rather than sampling uniformly from a population, one includes all individuals suffering the disease (since these are rare and valuable from the statistical point of view) plus a certain number of controls, that is, individuals not suffering the disease, sampled from the population. In our situation we have a "prevalence" of just one event among a "population" of up to 350 million dyads, which is definitly a very rare outcome. Case-control sampling in the context of relational event models has been proposed by [32]. Earlier [5] suggested sampling to approximate terms over the "support set" without specifying a concrete sampling scheme.

So, instead of summing over all pairs $(u', a') \in U_t \times A_t$ in the denominator of Eq. (2) we sum only over $(u', a') \in \textsc{sample}(U_t \times A_t; e)$, where the sample always includes the user-article pair on which the observed event e happened plus m further pairs that are uniformly selected at random from $U_t \times A_t$. In our concrete analysis we sample five non-event dyads for each event. Repeating the analysis 100 times revealed that the standard deviation of the parameters over the different samples is not larger than the estimated standard errors (see Table 1). To estimate model parameters and their standard errors from the sampled likelihood function we used the function `coxph` from the R-package **survival**[7] [30].

2.5 Explanatory variables (statistics)

The dyadic event rate $\lambda(u, a, x, t; G[E; t]; \theta^{(x)})$ has been specified as a function of statistics $s_j(u, a; G[E; t])$ in Eq. (1). Next we define the concrete statistics that we use in our model, illustrated in Fig. 2. For each of these statistics (except that

[7] https://CRAN.R-project.org/package=survival

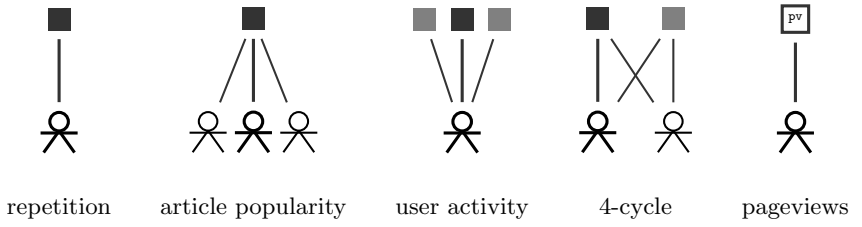

Fig. 2. The event rate on a given user-article pair (thick line, connecting dark gray nodes) is specified dependent on configurations encoding the following network effects. *Repetition*: dyadic event rates depend on past events on the same dyad. *Article popularity*: the event rate on dyad (u, a) depends on past events received by the same article a (but potentially initiated by other users). *User activity*: the event rate on dyad (u, a) depends on past events initiated by the same user u (but potentially directed towards other articles). *4-cycle*: the event rate on dyad (u, a) depends on past events forming a 3-path from u to a via different articles and users (see text). *Pageviews*: the event rate on dyad (u, a) depends on the number of pageviews on article a.

dependent on the number of pageviews) we have two variants, one dependent on past edit events and one dependent on past talk events. We note that statistics defined on past edit events are also used in the model explaining the rate of future talk events and vice versa. For a time point t, a user $u \in U_t$, an article $a \in A_t$, and an event type $x \in \{edit, talk\}$ we define the following statistics.

Edit repetition and talk repetition. If u contributed to a before (by editing or participation in discussion), it is likely that u has an interest in a's topic and thus is more likely to contribute again in the future. Such a hypothetical effect is expressed by the two statistics (for $x = edit$ and $x = talk$)

$$x.repetition(u, a; G[E; t]) = past.x(u, a, t) \ .$$

Article popularity. If a received many events from any user in the past, then a is popular in the community of contributing users and thus a is more likely to receive events at a higher rate in the future. Such a hypothetical "rich-get-richer" effect is expressed by the two in-degree statistics

$$x.popularity(u, a; G[E; t]) = \sum_{u' \in U_t} past.x(u', a, t) \ .$$

User activity. If u initiated many events towards any article in the past, then u is an active user and thus u is more likely to initiate events at a higher rate in the future. Such a hypothetical effect is expressed by the two out-degree statistics

$$x.activity(u, a; G[E; t]) = \sum_{a' \in A_t} past.x(u, a', t) \ .$$

Assortativity. The hypothetical effect that popular articles will receive even more contributions in the future could have negative consequences as less popular articles are in danger of being neglected. It is therefore crucial to understand who (if anyone) is more likely to start working on the less popular articles. We hypothesize that the more active users could fulfill this role. Thus we hypothesize to find negative assortativity between user activity and article popularity, expressing that highly active users are less likely to contribute to highly popular articles (all other things being equal) and therefore more likely to start working on less popular articles. Such a hypothetical effect can be tested by the two assortativity statistics. (Note that we hypothesize to estimate a negative parameter associated with them.)

$$x.assortativity(u, a; G[E; t]) = x.pop(u, a; G[E; t]) \cdot x.act(u, a; G[E; t]) \ .$$

4-cycles. Users might organize themselves into latent topics or disciplines. This would be expressed by 4-cycle effects (compare Fig. 2): if users u and u' already collaborated on an article a', then they are likely to be interested in the same topic. If, in addition, u' contributed to article a, then u is more likely to also contribute to a. Such effects can be tested by the two 4-cycle statistics:

$$x.4.cycle(u, a; G[E; t]) = \sum_{u' \in U_t \setminus \{u\}} \sum_{a' \in A_t \setminus \{a\}} \min[x(u, a', t), x(u', a', t), x(u', a, t)]$$

where we abbreviated $past.x$ by x.

Article pageviews. Last but not least we test against the alternative explanation that some or all of the above effects are just due to varying popularity of articles' topics. Recall that for each article a and each time period t we know the number of pageviews $pv(a, t)$, that is the number of times the article a has been requested during t by any Internet user. Time precision for the pageviews is one hour (while the time of edit events and talk events are known by the second). We interpret $pv(a, t)$ as a measure of general interest in, or popularity of, a's topic at time t. Since hourly pageviews are very volatile, we define a statistic by a smoothed moving average:

$$article.pageviews(u, a; G[E; t]) = \sum_{t' \leq t} pv(a, t') \cdot \exp\left(-(t - t') \cdot \frac{\ln(2)}{T_{1/2}}\right) \ ,$$

where the summation index t' runs over all intervals of one hour up to t and the *halflife* $T_{1/2}$ defines the value of the time difference $t - t'$ after which the influence of pageviews at t' is halved. In our analysis we set the halflife to one day, that is $T_{1/2} = 24$. This means that 1,000 pageviews today count as 500 pageviews tomorrow and as slightly less than one pageview in 10 days.

All statistics have been transformed by the mapping $x \mapsto \log(1 + x)$ and then standardized to mean zero and standard deviation equal to one. This standarization makes parameter sizes better comparable.

3 Results and discussion

Results of four estimated models (the models explaining edit events and talk events, with and without the pageview statistic) are shown in Table 1. Several of the parameters in the first model for edit events are as expected. For instance, we find a large positive parameter for *edit.repetition* implying that users are much more likely to contribute to articles they have edited before. We also find a rich-get-richer effect, so that articles that have already received many edits (by any user) are likely to receive more edits, by potentially different users, in the future (positive parameter of *edit.popularity*). We find a similar effect in the edit activity of users, such that users who have been more active in the past show a higher editing rate in the future (positive parameter of *edit.activity* in the edit model).

Table 1. Estimated parameters and standard errors (in brackets) for the models for edit events and talk events. The first two models are estimated on events from January 15, 2001 to January 1st, 2018. The last two models, which also include the pageview statistic, are estimated on events from May 1st, 2015 to January 1st, 2018.

	edit model	talk model	edit (pv)	talk (pv)
edit.repetition	5.018 (0.018)*	6.075 (0.075)*	5.196 (0.048)*	5.860 (0.195)*
talk.repetition	1.096 (0.020)*	5.803 (0.120)*	1.042 (0.047)*	4.936 (0.244)*
edit.popularity	0.924 (0.005)*	0.590 (0.022)*	0.679 (0.017)*	0.081 (0.069)
edit.activity	0.967 (0.004)*	-0.205 (0.017)*	1.622 (0.012)*	0.474 (0.048)*
talk.popularity	0.027 (0.004)*	0.465 (0.021)*	-0.057 (0.012)*	0.551 (0.058)*
talk.activity	-0.238 (0.004)*	1.328 (0.019)*	-0.291 (0.010)*	1.272 (0.057)*
edit.4.cycle	0.034 (0.004)*	-0.370 (0.018)*	-0.382 (0.011)*	-0.902 (0.047)*
talk.4.cycle	0.343 (0.005)*	1.120 (0.024)*	0.359 (0.012)*	1.133 (0.064)*
edit.assortativity	-0.258 (0.003)*	-0.026 (0.012)	-0.270 (0.008)*	0.115 (0.033)*
talk.assortativity	-0.088 (0.003)*	-0.655 (0.018)*	-0.090 (0.008)*	-0.743 (0.047)*
article.pageviews			0.736 (0.010)*	0.758 (0.043)*
Num. obs.	4,892,946	828,101	1,036,212	156,172
Num. events	815,722	138,036	172,552	26,027

* $p < 0.001$

Since users have a preference to edit already popular articles, there is the danger that the less popular articles get neglected and never accumulate a critical mass of user contributions. We hypothesized that active users might be the ones who start working on the less popular articles. The negative parameter of *edit.assortativity* supports this conjecture and indicates that highly active users are rather drawn to editing the less popular articles, all other things being equal. We also find a positive (albeit small) parameter for *edit.*4.*cycle* pointing to a tendency for local clustering into latent topics.

Turning to the influence of engagement in talk pages on editing, we make the observation that users contributing in discussion on any article are less likely to engage in editing (negative parameter of *talk.activity* in the edit model). This, together with the negative parameter of *edit.activity* in the talk model, indicates the presence of self-selected roles, such that some users engage mostly in editing and others rather in discussion. Besides this, we find similar structural effects in the talk model as in the edit model.

Controlling for the number of pageviews we find that this indicator for general popularity of an article's topic does have a positive effect on the rate of edit and talk events. Thus, internal popularity among contributing users partially reflects external popularity by the general public. However, the network effects within and across edit and talk events remain relatively stable, with some exceptions – most notably the effect of *edit*.4.*cycle* in the edit model. Thus we can conclude that network effects drive attention of contributing users beyond the impact of varying general popularity.

4 Conclusion

Attention by contributing users is one of the most crucial resources for the production of Wikipedia articles. In the absence of predefined centralized coordination mechanisms and the lack of financial rewards, we argue that emergent networks arising from task-oriented interaction explain, at least in part, the allocation of work to specific articles. In this paper, we specified and implemented relational event models to analyze the dynamic two-mode network in which users are connected by edit or talk events to the articles they contribute to. Importantly relational event models can analyze Wikipedia log data in the given granularity without the need to aggregate over time, users, or articles. We demonstrated that network effects are important drivers of attention received by articles.

We see several avenues for future work. Besides extending the analysis to a larger sample of article and including more characteristics of users and articles into the analysis, we could also exploit additional information available for individual edits. In this paper we modeled who is going to edit when which page, but treated edits as atomic events. The model in [20] instead modeled *how* users edit articles (for instance, whether they undo or redo contributions of others), given that they do upload a revision at a given point in time. Combining these two models could yield a more complete picture about coordination of collaboration in Wikipedia.

Acknowledgements

This work has been supported by Swiss National Science Foundation (FNS Project Nr. 100018_150126) and Deutsche Forschungsgemeinschaft (DFG Grant Nr. LE 2237/2-1).

References

1. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
2. Benkler, Y., Shaw, A., Hill, B.M.: Peer production: A form of collective intelligence. Cambridge, MA: MIT Press (2015)
3. Borgan, Ø., Goldstein, L., Langholz, B.: Methods for the analysis of sampled cohort data in the Cox proportional hazards model. The Annals of Statistics pp. 1749–1778 (1995)
4. Brandes, U., Lerner, J., Snijders, T.A.: Networks evolving step by step: Statistical analysis of dyadic event data. In: Proc. 2009 Intl. Conf. Advances in Social Network Analysis and Mining (ASONAM), pp. 200–205. IEEE (2009)
5. Butts, C.T.: A relational event framework for social action. Sociological Methodology **38**(1), 155–200 (2008)
6. Cox, D.: Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological) **34**(2), 87–22 (1972)
7. Foucault Welles, B., Vashevko, A., Bennett, N., Contractor, N.: Dynamic models of communication in an online friendship network. Communication Methods and Measures **8**(4), 223–243 (2014)
8. Georgescu, M., Pham, D.D., Kanhabua, N., Zerr, S., Siersdorfer, S., Nejdl, W.: Temporal summarization of event-related updates in Wikipedia. In: Proc. 22nd Intl. Conf. World Wide Web, pp. 281–284. ACM (2013)
9. Gómez, V., Kappen, H.J., Litvak, N., Kaltenbrunner, A.: A likelihood-based framework for the analysis of discussion threads. World Wide Web **16**(5-6), 645–675 (2013)
10. von Hippel, E., von Krogh, G.: Open source software and the "private-collective" innovation model: Issues for organization science. Organization science **14**(2), 209–223 (2003)
11. von Hippel, E., von Krogh, G.: Free revealing and the private-collective model for innovation incentives. R&D Management **36**(3), 295–306 (2006)
12. Kaltenbrunner, A., Laniado, D.: There is no deadline: Time evolution of Wikipedia discussions. In: Proc. 8th Annual Intl. Symp. Wikis and Open Collaboration. ACM (2012)
13. Karimi, F., Bohlin, L., Samoilenko, A., Rosvall, M., Lancichinetti, A.: Mapping bilateral information interests using the activity of Wikipedia editors. Palgrave Communications **1** (2015)
14. Keegan, B., Gergle, D., Contractor, N.: Do editors or articles drive collaboration? Multilevel statistical network analysis of Wikipedia coauthorship. In: Proc. 2012 Conf. Computer Supported Cooperative Work, pp. 427–436. ACM (2012)
15. Keegan, B., Gergle, D., Contractor, N.: Staying in the loop: Structure and dynamics of Wikipedia's breaking news collaborations. In: Proc. 8th Annual Intl. Symp. Wikis and Open Collaboration. ACM (2012)
16. Keegan, B.C., Lev, S., Arazy, O.: Analyzing organizational routines in online knowledge collaborations: A case for sequence analysis in cscw. In: Proc. 19th Conf. Computer-Supported Cooperative Work & Social Computing, pp. 1065–1079. ACM (2016)
17. Lawless, J.F.: Statistical Models and Methods for Lifetime Data. Wiley (2003)
18. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Computational social science. Science **323**(5915), 721–723 (2009)

19. Leenders, R.T.A., Contractor, N.S., DeChurch, L.A.: Once upon a time: Understanding team processes as relational event networks. Organizational Psychology Review **6**(1), 92–115 (2016)
20. Lerner, J., Lomi, A.: The Third Man: Hierarchy formation in Wikipedia. Applied Network Science **2**(1), 24 (2017)
21. Lerner, J., Lomi, A.: Knowledge categorization affects popularity and quality of Wikipedia articles. PloS one **13**(1), e0190,674 (2018)
22. Lerner, J., Tirole, J.: The open source movement: Key research questions. European economic review **45**(4), 819–826 (2001)
23. Maddock, J., Shaw, A., Gergle, D.: Talking about talk: Coordination in large online communities. In: Proc. 2017 CHI Conf. Human Factors in Computing Systems, pp. 1869–1876. ACM (2017)
24. Moat, H.S., Curme, C., Avakian, A., Kenett, D.Y., Stanley, H.E., Preis, T.: Quantifying wikipedia usage patterns before stock market moves. Scientific reports **3**, 1801 (2013)
25. Pilny, A., Schecter, A., Poole, M.S., Contractor, N.: An illustration of the relational event model to analyze group interaction processes. Group Dynamics: Theory, Research, and Practice **20**(3), 181 (2016)
26. Romero, D.M., Huttenlocher, D., Kleinberg, J.M.: Coordination and efficiency in decentralized collaboration. In: ICWSM, pp. 367–376 (2015)
27. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental study of inequality and unpredictability in an artificial cultural market. Science **311**(5762), 854–856 (2006)
28. Schecter, A., Contractor, N.: Understanding and assessing collaborative processes through relational events. In: Innovative Assessment of Collaboration, pp. 223–231. Springer (2017)
29. Stroud, N.J.: Attention as a valuable resource. Political Communication **34**(3), 479–489 (2017)
30. Therneau, T.M., Grambsch, P.M.: Modeling survival data: extending the Cox model. Springer Science & Business Media (2013)
31. Viegas, F.B., Wattenberg, M., Kriss, J., Van Ham, F.: Talk before you type: Coordination in Wikipedia. In: Proc. 40th Annual Hawaii Intl. Conf. System Sciences (HICSS'07), p. 78a. IEEE (2007)
32. Vu, D., Pattison, P., Robins, G.: Relational event models for social learning in moocs. Social Networks **43**, 121–135 (2015)
33. Webster, J.G.: The marketplace of attention: How audiences take shape in a digital age. Mit Press (2014)
34. Wu, T.: The attention merchants: The epic scramble to get inside our heads. Vintage (2017)
35. Yasseri, T., Sumi, R., Kertész, J.: Circadian patterns of wikipedia editorial activity: A demographic analysis. PloS one **7**(1), e30,091 (2012)
36. Zhu, M., Huang, Y., Contractor, N.S.: Motivations for self-assembling into project teams. Social Networks **35**(2), 251–264 (2013)