Discussion Paper No. 14-132

Spillovers in Networks of User Generated Content

Pseudo-Experimental Evidence on Wikipedia

Michael Kummer

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 14-132

Spillovers in Networks of User Generated Content

Pseudo-Experimental Evidence on Wikipedia

Michael Kummer

Spillovers in Networks of User Generated Content*

Pseudo-Experimental Evidence on Wikipedia

Michael E. Kummer[1]

December 29, 2014

Abstract

I quantify spillovers of attention in a network of content pages, which is challenging, because such networks form endogenously. I exploit exogenous variation in the article network of German Wikipedia to circumvent this problem. Wikipedia prominently advertises one featured article on its main site every day, which increases viewership of the advertised article. Shifts in the viewership of adjacent articles are due to their link from the treated article. Through this approach I isolate how the link network causally influences users' search and contribution behavior.

I use a difference-in-differences analysis to estimate how attention spills to neighbors through the transient shock of advertisement. I further develop an extended peer effects model which relaxes the requirement of an exogenously given network. This model enables the estimation of the underlying spillover. Advertisements affect neighboring articles substantially: Their viewership increases by almost 70 percent. This, in turn, translates to increased editing activity. Attention is the driving mechanism behind views and short edits. Both outcomes are related to the order of links, while more substantial edits are not.

Keywords: Social Media, Information, Knowledge, Spillovers, Networks, Natural Experiment
JEL Classification Numbers: L17, D62, D85, D29

* I thank the IIIrd ICT Conference (Munich) for selecting an early version as runner up in the "Best PhD Paper Award." I benefitted from discussions with Irene Bertschek, Christoph Breunig, Luis Cabral, Jörg Claussen, Ulrike Cress, Habiba Djebbari, Neil Gandal, Avi Goldfarb, Sanjeev Goyal, Shane Greenstein, Thorsten Grohsjean, Stephen Kastoryano, Maximilian Kasy, Francois Laisney, Jose Luis Moraga-Gonzalez, Kathleen Nosal, Gal Oestreicher-Singer, Martin Peitz, Marianne Saam, Philipp Schmidt-Dengler, Patrick Schulte, Jan-Peter Siedlarek, Olga Slivko, Yossi Spiegel, André Stenzel, Joel Waldfogel, Michael Ward, Andrea Weber and Michael X. Zhang. I thank participants in several seminars at ZEW and U Mannheim, as well as JKU Linz, MIT, Northeastern U, U Kaiserslautern and WU Wien and of the following conferences: 37th NBER Summer Institute (Boston), 5th SEARLE "Internet Search and Innovation" (Chicago), 12th IIOC (Chicago), 11^{th} Media Economics (Tel Aviv), EEA-ESEM 2014 (Toulouse), 7th ICT (ParisTech), IIIrd ICTCM (Munich), 6th "Natural Experiments and Controlled Field Studies" (Ohlstadt), EARIE 2013 (Evora) and 11^{th} ICT (Mannheim). I am grateful to the Wikimedia Foundation and Frédéric Schütz for access to the Wikipedia data. I thank Thorsten Doherr and Manfred Knobloch for support with data processing and to George Giorgidze for help with DSH. James Binfield, Timo Schäfer and Daniel Bergman provided outstanding research assistance. Financial support from the WissenschaftsCampus Tübingen is gratefully acknowledged.

[1] Centre for European Economic Research (ZEW), Mannheim and University of Mannheim; L7 1; 68181 Mannheim, Germany; Kummer@zew.de.

1 Introduction

Carefully placed links in scientific citation networks could contribute to the accelerated narrowing of structural holes in knowledge. Specifically, citations might divert researchers' attention, and such diversion might cause them to contribute content, effort, or ideas to the cited field. Applied in knowledge-based peer production, such a mechanism may furnish profound policy implications.

In this paper, I show how links channel users toward viewing adjacent articles and contributing additional content in German Wikipedia. I use exogenous and transient shocks to attention to identify this effect. Three aspects make Wikipedia a relevant object of study: first, it is peer produced; second, the "clickable blue words" in Wikipedia's articles are highlighted hyperlinks which form a citation network; and third, Wikipedia has almost completely taken over the market for encyclopaedic information from previous incumbents, such as Encyclopaedia Britannica.

The causal effect of citations on attention is difficult to measure, but I apply a careful treatment-control design to overcome the methodological challenges. Correlations between the attention to an article and its citations typically abound in any relevant citation network - be it citations among patents, scientific papers or blogs. However, such correlations do not reflect the causal effect of citations on attention, because both variables co-evolve and mutually reinforce each other. I can circumvent this endogeneity problem by exploiting an institutional feature of German Wikipedia, "Today's featured article". This is a specific advertisement on Wikipedia's front page, where a single article is advertised for exactly one day. It substantially increases an article's viewership and is exogenous to "a normal day's" content production.

The basic idea of my research approach can be imagined as "throwing stones into a pond and tracing out the ripples". I document how the transient shock from advertisement spills on to the neighbors by using a difference-in-differences estimator. The first difference in the pseudo-experiment is between the days before advertisement and the day of advertisement. The second difference is between neighbors of "*Today's* featured article" and neighbors of a featured article to be advertised in the *future*.

I set up 93 such pseudo-experiments using a special database that combines data on Wikipedia's revisions, the link structure and page views. I extract data which contain four weeks of daily information on 186 featured articles and all their direct and indirect neighbors (more than 15,500 pages and 900,000 observations). For 93 featured articles I observe the day, when they were advertised as "*Today's* featured article". Advertisement triggers more than 4,000 additional views. This is a 40-fold increase over average viewership ("the stone"). Roughly 70 percent of these readers click on a link, which leads to a corresponding increase in views and contributions on neighboring pages ("the ripple").

Moreover, I develop an extended version of the empirical peer-effects model by

Bramoullé et al. (2009) to estimate the structural spillover effect. I exploit the shocks to relax the requirement of an exogenous (to the variables of interest) network structure, and merely require it to be stable over short periods of time. The combination with difference-in-differences allows me to uncover the underlying "average day" attention spillovers. I derive an interval estimator which can be computed even without information on the link structure. These methods apply even if identification through partial overlaps in the network structure fails. For German Wikipedia, ten more average views on the neighboring pages turn out to generate 1.9 to 2.41 more views.

Having analyzed the attention spillover, I quantify how more attention generates new content. A thousand views generate one edit and I find that content generation is not distributed equally between articles. While attention spillovers do not depend on the length of the link's target, content generation *does*. Moreover, analyzing the content changes' persistence suggests that attention-driven contributions to neighbors tend to be superficial. These context-specific findings should encourage using my approach on other citation networks. The large effects for attention suggest that policy makers should consider using citations to channel attention in similar contexts such as R&D. Also companies seeking to document firm-specific knowledge in a wiki would offer a fruitful environment for application.

In sum, my analysis of the link network's causal effects reveals a robust pattern: links channel attention which, in turn, drives contributions. Beyond quantifying the relevance of citations, this insight can help leveraging the potential of "collaborative peer production." This is of great economic interest to both society and firms, since the new production mode might drastically reduce the cost of producing or managing knowledge. Moreover, I provide a showcase of exploiting exogenous variation in a network to estimate how nodes affect each other mutually. Whenever network and outcome are suspected to co-evolve, be it for citations, banking, trade or in social networks, this should be the method of choice.

I discuss the contribution to the literature in Section 2 and identification in Section 3. Section 4 discusses the data collection and presents basic descriptive results. Section 5 extends the linear-in-means peer effects model to allow for local treatments in the network and estimates the underlying average spillover. Detailed derivations of the estimator and the bounds are in Appendix E. Section 6 scrutinizes content generation: I ask which articles will more likely receive edits and show how much the content is improved. I discuss findings, limitations and avenues for further research in Section 7. Section 8 concludes. Further appendices contain summary statistics, robustness checks and additional figures.

2 Literature

The success of Open Source Software production and Wikipedia fundamentally challenged Coase's (1937) insights, which held that production should either be organized in a free market if market frictions are low, or in a firm if they are high. The new coordinating principle, by which large numbers of people distribute small modules of the total workload via the web is referred to as commons-based peer production (Benkler 2002 and 2006). The extraordinary past achievements of this production mode illustrate the deep impact its emergence might have on the economic process and even society as a whole.

This paper contributes to the literature in three ways. First, I document how attention spills over the link network in a relevant setting of peer production - the German Wikipedia. Second, I quantify how attention is converted into action (contributing content). Finally, I analyze the heterogeneities in the spillovers in the network. In what follows, I discuss the streams of the literature that each of these contributions add to.

I add to previous research, which analyzed how networks generate externalities and influence real world outcomes.[1] Economists have asked how social networks influence economic real world outcomes for (at least) two reasons: First, it is important to understand how a network's structure affects individuals' outcomes and to quantify the resulting overall value of a network and its links. Second, it matters whether peers mutually influence each others' outcomes, be it positively or negatively. Among other things, such influences may lead to important multiplier effects of interventions. My paper quantifies the causal effect of the average attention of a focal articles's neighbors on the attention of the focal article. Previous research on such network spillovers has struggled with the following empirical problems: The outcome variable might itself drive network position, thus giving rise to the classic endogeneity problem.[2] These problems have lead researchers to adopt an "intellectually unsatisfying" division of research which focused either on network formation or the effects of the network (Graham (2015.)). However, until today researchers avoid analyzing them jointly (cf. Jackson and Zenou (2013)). Moreover, the reflection problem laid out by Manski (1993) applies, since nodes influence each other like peers (Bramoullé et al. (2009)). This paper shows how both problems can be circumvented by exploiting local exogenous treatments of single nodes in Wikipedia's article network.

A second contribution to the literature is the econometric approach to quantifying attention spillovers between Wikipedia articles. My formal framework combines existing approaches and extends them in a novel way, because I incorporate local treatments

[1] There is a well documented correlation between a node's position in a network and the outcomes of interest (Fershtman and Gandal (2011), Claussen et al. (2012) or Kummer et al. (2012)). Moreover, several papers have shown how peers influence each other's outcomes in education and used instruments based on partial overlaps in the network structure to solve the reflection problem described below. (Calvó-Armengol et al. (2009), De Giorgi et al. (2010), etc.)

[2] An example of how this can emerge naturally in a similar setting can be found in models of linking among blogs. (cf. Mayzlin and Yoganarasimhan (2012) or Dellarocas et al. (2013))

into a model of peer effects. Yet, instead of focusing on the effect of treatment I focus on the *spillovers* of these treatments and use them as sources of exogenous variation in the *attention* to such articles. Moreover, I use the fact that exogenous treatment sometimes affects only a single node. Such local treatments are analogous to the Partial Population Treatment that Moffitt (2001) suggested for the analysis of peer effects - not in the context of network analysis - to solve the reflection problem that is identified in Manski (1993).[3] There is also a close relationship to studies that add a higher layer of randomization, which allows the computation of *indirect* treatment effects.[4] An example is Crépon et al. (2013), who randomize over cities and vary the treatment intensity to study whether labor market programs have a negative impact on the non-eligible. Studies that use exogenous local shocks to single individuals could be called "Mini Population Treatments" and this idea can be found in recent studies that use network information (Aral and Walker (2011), Banerjee et al. (2012), Carmi et al. (2012)).

Following the analysis of attention spillovers, I analyze how attention translates into content generation in a second step. Note the double function of the indirect treatment effects, which serve both as dependent variable in the first, and as independent variable in the second step.[5] I find a conversion rate of 1000 clicks for 1 edit. These findings highlight the need of adding an important extra ingredient to modularity and strong leadership (Benkler and Nissenbaum (2006), Lerner and Tirole (2002)), to guarantee the success of peer production: If the individuals contribute infrequently, a high overall frequency of visits is necessary. This reaffirms the potential of new information and communication technologies (ICTs) to enable peer production through their ability to drastically reduce coordination costs. These findings shed light onto the question how attention influences the decision to contribute to a public good. This question is relatively novel and only few papers have previously analyzed it. Several papers show that attention through blogs or reviews, even if negative, can be positively related to purchase and investment decisions (Barber and Odean (2008), Berger et al. (2010), Hu et al. (2013)). However, it is typically impossible to measure the amount of attention generated by the publicity and how it is converted to action.[6]

[3] Dahl et al. (2012) provide an example of such an experiment. An alternative approach is to exogenously vary the composition of peer groups: Zhang and Zhu (2011) uses the fact almost all Chinese Wikipedia users in mainland China were blocked by the government's "Chinese fire wall", to measure the effect on the incentives to contribute. Also disasters or fatal accidents are frequently used in similar settings. (Sacerdote (2001), Imberman et al. (2009)), Ashenfelter and Greenstone (2004)). Keegan et al. (2013), who analyze the structure and dynamics of Wikipedia's coverage of breaking news events.

[4] When social effects or spillovers are present, a violation of Stable Unit Treatment Value Assumption (SUTVA) compromises the validity of the control group (Ferracci et al. (2012)). Depending on the application, a second layer (classrooms, villages, districts etc.) can remedy the issue. (Miguel and Kremer (2003), Angelucci and De Giorgi (2009), Kuhn et al. (2011) and many more).

[5] The effect of advertisement on the neighbors is the coefficient of interest when quantifying attention spillovers. Yet, these spillovers incidentally generated exogeneous variation in the attention to articles (the neighbors), which allows estimating the effect of attention on edits.

[6] Altruism and social image concerns are important drivers of voluntary provision (non-monetary)

Finally this paper contributes by analyzing whether attention spills uniformly or whether there are large heterogeneities. I analyze the mediating factors of attention spillovers and subsequent content generation. Carmi et al. (2012) pioneering work analyzes what determines whether spillovers take place or not. They find that the network structure does well in predicting spillovers on Amazon's recommendation network.[7] I distinguish articles by their length, their connections, by the link's position and by how closely they are linked to the shocked articles. Only few other studies have analyzed which items receive collective attention. (Hoffman and Ocasio (2001), Wu and Huberman (2007)).[8] I contribute by analyzing what user choose, when presented with several options for a click and the subsequent conversion of awareness to making a voluntary (non-monetary) contribution to a public good.

3 Identification Strategy

In this section I outline the basic intuition of my estimation approach and provide background information on "Today's featured article" (Subsections 3.1 and 3.2). I then discuss the assumptions for identifying the treatment effects (Subsection 3.3) and estimate them in subsection 3.4. Finally, I discuss the identification and estimation of the underlying spillover in the separate section (Section 5) on attention spillovers. This is because it takes several steps to obtain this coefficient from the treatment effects. Specifically, I extend the linear-in-means model of peer effects to include local exogenous shocks.

3.1 Intuition - Throwing Stones into a Pond

The basic idea of my research approach can be imagined as "throwing stones into a pond and tracing out the ripples." The schematic representation in Figure 1 shows how the data are structured. "The pond" is the network of Wikipedia articles. In this network every article is a node and is represented by a circle with a letter inside. Thus, each circle represents a different article in the German Wikipedia. Articles are connected to each other via links, which are visible on Wikipedia as highlighted blue text. Clicking on such

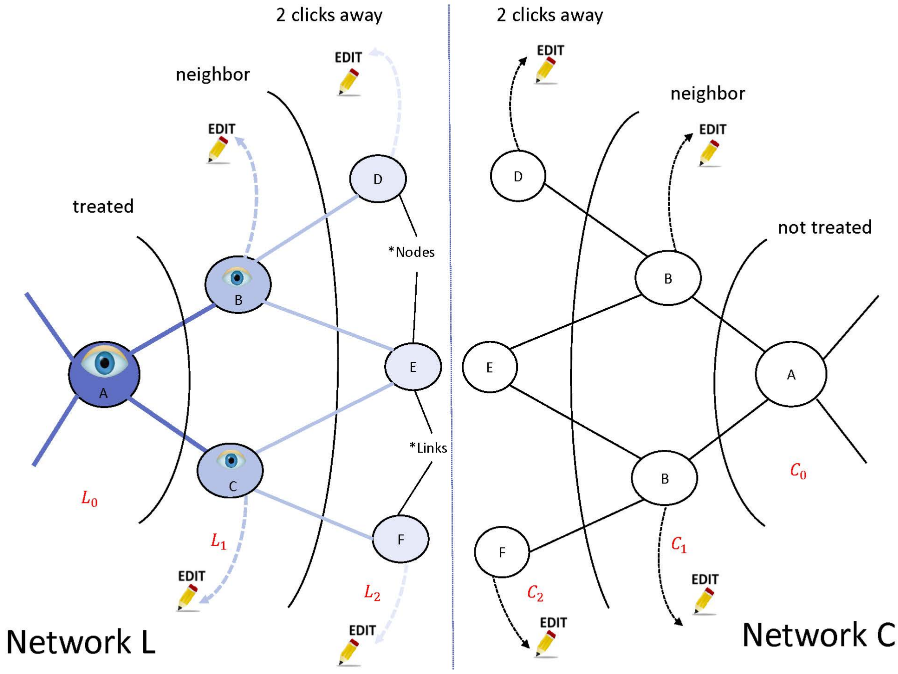
of a public good in offline contexts (Carpenter and Myers (2010)). Social effects and attention to the individual contribution also matter in peer productivity (Shang and Croson (2009), Huberman et al. (2009)). Yet, studies that precisely quantify how attention converts to contributions and that disentangle this effect from the other relevant drivers of contributions are rare.

[7] Carmi et al. (2012) analyze the effect of the external shocks of recommendations by Oprah Winfrey on the product network of books on Amazon. They find that a recommendation not only triggers a spike in sales of the recommended book but also of books adjacent in Amazon's recommendation network. Unfortunately, their findings are challenged by the fact that spillovers are also an important driver of Amazon's algorithm that places and sorts the links.

[8] Viral Marketing studies are concerned with the diffusion of information in a social network, i.e. mediated by social propagation, rather than repeated individual decisions.(e.g. Aral and Walker (2011), Ho and Dempsey (2010), Hinz et al. (2011))

a blue link forwards the reader to the next article and these links form the edges of my directed network. In the figure below links are represented by a line between two nodes.

Figure 1: Schematic representation of a local treatment, which affects only one of the two subnetworks and there only a single node directly.



NOTES: The figure illustrates the structure of the data. Wikipedia articles are the nodes of the network. Each circle with a letter inside represents a different article in the German Wikipedia. The eye icons represent attention, while the pencils illustrate a decision to contribute an edit. Articles are connected to each other via links, which are represented as lines. The design of this paper uses the fact that certain nodes were affected by a large and exogenous increase of attention, and that it is known to the researcher when the pseudo-experiment occurred. In this setting this is represented by the two subnetworks L and C. Both, nodes in L_0 and C_0, could be hit by a disaster (or are featured articles). Hence they are eligible for treatment. Yet, only one is actually hit (or becomes "*today's* featured article") at any given day. The coloring illustrates the effect of such a large local shock on Wikipedia, which affects only subnetwork L. The shocked node is colored in dark blue, the direct neighbors are colored in light blue and so on. If we observe a valid second network from which it is possible to infer what the outcomes would have been if no treatment had taken place, we can use these outcomes for comparing the size of the outcomes layer by layer. In general the network may be directed or undirected (Wikipedia articles are directed). The figure draws on a representation in a working paper on network formation by Claussen, Engelstaetter and Ward.

The "stone in the pond" is generated by Wikipedia's "Today's featured article," which is a mechanism that very prominently advertises a single article on Wikipedia's front page. This advertisement generates a large temporary increase in article traffic. I will sometimes refer to "Today's featured article" as "the advertised article" and I shall discuss the details and institutional aspects of "Today's featured article" in the next subsection.

An important aspect of my identification strategy is that it requires the observation of two disconnected subnetworks at the same time. Hence, I show two networks, L and C, which face each other. They are both networks around start articles denoted by ℓ_0 and c_0. "Today's featured article" belongs to Network L. Network C consists of a set of similar articles, which are not affected by "Today's featured article" and provide valid control observations. The similarity is ensured because both start nodes, ℓ_0 and c_0, are featured high-quality articles and are both eligible to become "Today's featured article."

In fact, the articles in c_0 even *were* advertised at a later day. The details of how the pages in network C are obtained will be discussed in Section 4.2. I shall continue to use L and C to denote the two networks in all derivations that follow.

Articles that receive a direct link from a start article are direct neighbors. In network L they form the set L_1 and a specific node from that set is denoted as ℓ_1.[9] The set of indirect neighbors, which are two clicks away in the network L, forms L_2 and so on.[10] Analogously the set C_1 are the direct neighbors of the non-advertised start article c_0 in network C, and C_2 are the indirect neighbors. The coloring in Figure 1 illustrates the mechanism of local exogenous shocks ("the stone in a pond"). The shocked node is "Today's featured article." It receives a lot of additional attention and is hence colored in dark blue. The direct neighbors ("the nearest ripple") receive the direct spillover and are colored in light blue, indirect neighbors receive much less and so on.

In more general network terminology, the design of this paper uses the following facts: First, certain nodes were affected by a large and exogenous increase of attention, where the exogeneity is with respect to the "normal" production and linking process. Second, it is known (ex-post) to the researcher when exactly the pseudo-experiment occurred. Moreover, since the link structure is also known, it is possible to observe what happens to the directly or indirectly neighboring nodes. As in a pond, we would expect the largest effect on the directly hit node and a decreasing amount of additional attention the further away an article is from the center.

In a typical network in which the outcome of the individual nodes depends on the outcome of their neighbors we would observe many correlations and cross influences. Absent local exogenous shocks, it would be difficult to discern where they originate from. Moreover, it would be hard to ascertain that no underlying factors and unobserved background factors merely affect similar nodes in similar ways. Identification of the spillover will require observing a valid "comparison network," C, to infer the counterfactual outcome of treatment layer by layer. I provide more information about how the comparison groups were obtained in Section 4.

3.2 Background - Links and "Today's Featured Article"

Before discussing the identifying assumptions and threats to identification, I provide background information about three relevant aspects of Wikipedia and the Wiki technology. These aspects are, the link network of articles, featured articles and "*Today's* featured article" which is at the heart of my design.

The first important aspect of Wikipedia for this study are the "blue word" hyperlinks between articles. Links are typically not placed randomly but Wikipedia's guidelines

[9]While the set L_0 consists only of one node ($L_0 = \{\ell_0\}$), set L_1 consists of multiple nodes.

[10]Indirect neighbors are defined as receiving at least one link from a node in set L_1 without themselves being in L_1. Hence the shortest path from the start node to an indirect neighbor is via two clicks.

require that they to refer to relevant additional information. Within those guidelines users still have the freedom to choose which links to place (and sometimes argue, whether a link should be present or not). Moreover links are only placed if users choose to include the corresponding term into the referring article. Similarly highly relevant articles are typically longer than the average article. This implies a positive correlation between the relevance of an article and both the number of links to the article and its length. Like researchers following up on a citation in a paper, users can click on the links to view the other article. While these are arguably not the same processes they share the notion of seeking more information.[11] Throughout the paper I interpret these links between articles as the citation network. From a network perspective, articles receive traffic because of these clickable links. Many frequented neighbors, generate more "redirected" traffic, which is the object I study. The challenge for identification is as follows: having more or fewer links and thus being in the neighborhood of other frequently linked pages are endogenous long-run outcomes. They both depend on where users pay attention and where they generate content.[12]

The second important institutional aspect are "featured articles." "Featured" is Wikipedia's top quality status for articles which cover all the relevant information in a particularly well-written and well-structured way. The community of Wikipedia editors awards the status in a well-defined procedure which involves a nomination and a review period. Some of these articles cover very important topics. However, since all editors are volunteers, many articles cover special interest topics that are not very frequently consulted. For example, warships and battles are clearly overrepresented. On January 1 in 2008, the German Wikipedia had a 1,241 featured articles and 668 were promoted to this status in the period of observation (2008-2010).

"*Today's* featured article" is the specific "featured articles," which is selected to be advertised on the front page for 24 hours ("today"). There is only one such article every day, and, as I will show later, such advertisements generate a lot of attention (which is measured as additional page views). Articles can be nominated ahead of time for being advertised on a specific day. This is necessary, because the pages are fitted to a designated box on the front page. Like all editing and most of the maintainance of Wikipedia, "Today's featured articles" are managed by a dedicated group of volunteers. These are up to ten individuals who collaborate to ensure that every day a new featured article is advertised. During the period of observation, the decision which pages to advertise next week was made by the end of the previous week. For conflicting nominations, a ballot was cast. Every editor could participate, but, in practice, the relevant administrative pages are hard to find and not many users participated.[13] Nowadays the second category

[11]Clicking the link follows up on less scientific information and involves fewer frictions than looking up a paper. Yet, in scientific papers the readers' incentives to follow up on a citation may be higher.

[12]Likewise, intertemporal variation in article traffic likely affects neighboring nodes in similar ways.

[13]Articles were infrequently nominated more than two weeks in advance. Up to 50% of the slots

of 3,713 "good articles" are also eligible.

My identification strategy is based on the fact that the featured articles, which were not advertised "today" (but later), are a valid control observation. The local effect is identified if the click through from "Today's featured article" to its neighbors is unaffected by swapping the timing of the advertisements. I emphasize that identification is not based on observing a random article or on the initial interest triggered by the advertised article. I will clarify my identifying assumptions further in the next section.

3.3 Identifying Assumptions for the Treatment Effects

I apply the control-treatment notation from impact evaluations Angrist and Pischke (2008) to describe my difference-in-differences (DiD) estimation. This clarifies the identifying assumptions and highlights similarities to the *partial population treatment* (cf. Moffitt (2001)). The terminology and notation are inspired by Kuhn et al. (2011). Since especially the direct treatment effect is standard (cf. Angrist and Pischke (2008)), a more detailed account of these assumptions is relegated to an online appendix.

Direct Effect of Treatment: Let the outcome of interest (y) be page views and consider a node in network $i \in \{L, C\}$ in period t. The subscript ℓ_t denotes nodes in the treated subnetwork in period t and c_t are nodes in the untreated subnetwork in period t. The direct treatment effect is defined as:

$$\text{(1)} \qquad \mathbf{E}[y^1_{\ell_0,t}|d_{\ell_0,t} = 1] - \mathbf{E}[y^0_{\ell_0,t}|d_{\ell_0,t} = 1]$$

where $d_{i,t}$ indicates if node i itself was directly treated or not. Superscript 1 denotes the outcome of a treated observation. The counterfactual outcome $\mathbf{E}[y^0_{\ell_0,t}|d_{\ell_0,t} = 1]$ is estimated from eligible articles in the untreated subpopulation.

Assumption Direct Treatment Effect-DiD:

$$\text{(2)} \qquad \begin{aligned} \mathbf{E}[y^0_{\ell_0,t}|d_{\ell_0,t} = 1] &= \mathbf{E}[y^0_{\ell_0,t-1}|d_{\ell_0,t-1} = 0] + \\ &+ \{\mathbf{E}[y^0_{c_0,t}|d_{c_0,t} = 0] - \mathbf{E}[y^0_{c_0,t-1}|d_{c_0,t-1} = 0]\} \end{aligned}$$

I estimate the counterfactual with last period's value *plus* the comparison group's rate of change. The identifying assumption is that treated observation, $y^0_{\ell_0}$, and the control $y^0_{c_0}$ grow at similar rates and are affected similarly by any dynamics that affect the entire Wikipedia (weekday dynamics etc.). The same applies to the indirect treatment effect:

"Indirect Treatment Effects": An ITE is slightly less standard. It measures the externality effect of eligible articles' treatment on the outcomes of the non-eligible.

stayed empty for the upcoming week and advertisement could be a surprise even for key authors (see http://de.wikipedia.org/wiki/Wikipedia_Diskussion:Hauptseite/Artikel_des_Tages#.28Haupt-.29Autoren_des_vorgeschlagenen_AdT_informieren for an example).

Neighbors are not themselves treated. I write $d^{\mathbf{0}}_{i,t}$ as shorthand for $d_{i,t} = 0$. For direct neighbors:

$$ITE_1 = \mathbf{E}[y^{\mathbf{1}}_{\ell_1,t}|D^{\mathbf{1}}_{\ell_1,t}, d^{\mathbf{0}}_{i,t}] - \mathbf{E}[y^{\mathbf{0}}_{\ell_1,t}|D^{\mathbf{1}}_{\ell_1,t}, d^{\mathbf{0}}_{i,t}] \tag{3}$$

where $D^{\mathbf{1}}_{\ell_1,t}$ takes the value 1 if there exists a treated node with a shortest distance of just 1 click to node ℓ_1. Like before $\mathbf{E}[y^{\mathbf{0}}_{\ell_1,t}|D^{\mathbf{1}}_{\ell_1,t}, d^{\mathbf{0}}_{i,t}]$ is counterfactual. To estimate it with difference-in-differences requires a set of comparable but untreated subpopulations (e.g. villages, classrooms, or here, subnetworks) and information about which individuals/nodes are eligible for treatment and which are not. The assumption is written as:

Assumption ITE-DiD:

$$\begin{aligned} \mathbf{E}[y^{\mathbf{0}}_{\ell_1,t}|D^{\mathbf{1}}_{\ell_1,t}, d^{\mathbf{0}}_{i,t}] &= \mathbf{E}[y^{\mathbf{0}}_{\ell_1,t-1}|D^{\mathbf{0}}_{\ell_1,t-1}, d^{\mathbf{0}}_{i,t-1}] + \\ &+ \{\mathbf{E}[y^{\mathbf{0}}_{c_1,t}|D^{\mathbf{0}}_{c_1,t}, d^{\mathbf{0}}_{i,t}] - \mathbf{E}[y^{\mathbf{0}}_{c_1,t-1}|D^{\mathbf{0}}_{c_1,t-1}, d^{\mathbf{0}}_{i,t-1}]\} \end{aligned} \tag{4}$$

I exploited the network's layers to define the ITE_1 (direct neighbors). We could also estimate the effect for indirect neighbors (ITE_2), simply replacing ℓ_1 with ℓ_2 etc.

Given this setup, I now clarify my assumptions for identification. The effect of interest is how views of article i depend on the views in the neighborhood. The essential assumptions can be phrased as the following question: "Had we swapped the timing of the advertisement, would there have been a peak and would the click through have been the same?" I emphasize that identification is based on the *click through*, but not on the shock being completely independent of other page characteristics. Specifically, identification does not require picking a representative article. In fact, all "featured articles" are better and more detailed than the average Wikipedia article. They also have more activity and more links. With theses assumptions, a valid counterfactual is given if the *set of neighboring articles* is on average comparable across treated and comparison neighborhoods. If the click through is further assumed to be the same as for advertising a random article, my setup identifies the average effect. A threat to identification might be the preference to advertise articles which are in some relationship to the current date, especially if e.g. anniversaries affected the click through rate. This concern can be addressed in a robustness check which tests this for advertisements on anniversaries.

The most important threat to identification is the fact that it is in principle possible to anticipate the advertisements. First, I discussed above, that some editors know about the advertisement in advance, and nodes are even nominated. However, the process was less structured in the period of observation. The advertisement can come as a surprise even for "key authors" of the article and often 50% of the slots stay empty for the upcoming week.[14] These slots are arguably filled based on "random" selection. Even so, a threat

[14]It is somewhat difficult to verify this ex-post, but a common theme in several entries in the forums,

to identification may be, if some links are placed in anticipation of the advertisement. I tackle this threat in two ways. First, I only include articles that had the link more than seven days before the advertisement of their neighbor. Second, I run a robustness checks, where the shocks come from sudden onset catastrophes, such as plane accidents, which are arguably beyond the control of the platform owners.

3.4 Estimation of Difference-in-Differences

I estimate the impact of local treatment with reduced form regressions for the treated pages and their neighbors. These regressions are similar to the analysis in Carmi et al. (2012). The outcomes of interest (y) are page views and edits. Treated and neighboring pages are regressed separately and I compare them to their analogue in the control group (L_0 to C_0, L_1 to C_1). I denote all reduced form coefficients by ϕ. Furthermore, I define "treatment" for each set of pages as direct or indirect treatment effect as in the previous section.[15] Here, s is time normalized around the day of treatment (day 0). Hence s runs from -14 to 14 and λ_s denotes one of the 29 corresponding time dummies. This results in the following system of fixed effect regression equations, based on 29 time dummies and 29 interaction terms which indicate a treated (or "shocked") observation in period t:

L_0.) The difference-in-differences specification at level L_0 is given by:[16]

$$y_{it} = \phi_i^{L_0} + \sum_{s \in S} \phi_{1,s}^{L_0} \lambda_s + \sum_{s \in S} \phi_{2,s}^{L_0} (\lambda_s * treat_{L_0,i}) + \xi_{it} \tag{5}$$

...$treat_{L0}$: treatment on the very page; $S = \{-14, ..., 14\}$

L_1.) At level L_1 ($treat_{L1}$ means the shock is 1 click away), the regression equation is:

$$y_{it} = \phi_i^{L_1} + \sum_{s \in S} \phi_{1,s}^{L_1} \lambda_s + \sum_{s \in S} \phi_{2,s}^{L_1} (\lambda_s * treat_{L_1,i}) + \xi_{it} \tag{6}$$

In words, I run the same simple difference-in-differences dummy regressions with 29 periods on L_0 and L_1. The outcome y can be page views and edits.[17] $treat_{L0,i}$ indicates a featured article that is to be advertised on Wikipedia's frontpage. Analogously, $treat_{L1,i}$ indicates the neighbors of such shocked pages. The cross terms capture whether treatment

such as http://de.wikipedia.org/wiki/Wikipedia_Diskussion:Hauptseite/Artikel_des_Tages/Archiv1#Mitteilung_auf_Diskussiosseite or http://de.wikipedia.org/wiki/Wikipedia_Diskussion:Hauptseite/Artikel_des_Tages/Archiv1#Akutes_Sommerloch_bei_den_Augustvorschl.C3.A4gen.

[15] The dummy in the regression for the neighbors (sets L_1 and C_1) takes the value 1, not if the node was itself treated, but if the corresponding start node (ℓ_0) was treated in t (and 0 otherwise).

[16] The specifications I use are fairly standard "regression difference in differences" similar to Jacobson et al. (1993) or as described in Angrist and Pischke (2008).

[17] Analogous regressions can also be run with other outcomes, such as new authors etc.

has occurred at a given point in time or not. Hence, the coefficients of treated observations are expected to be 0 before treatment occurs (for $s < 0$) and may be positive for the periods after the treatment (for $s \geq 0$). The coefficients in ϕ_1 capture the effect of the time dummies. The $ITEs$ from the previous subsection are captured by the ϕ_2 coefficient that corresponds to day 0 in the regressions above. $\phi_{2,0}^{L_1}$ measures the ITE_1, which corresponds to L_1 and $\phi_{2,0}^{L_0}$ represents the direct treatment effect for L_0.[18] Including page fixed effects accounts for page heterogeneity and the 29 time dummies control for time-specific activity patterns in Wikipedia.

This procedure is crude because it does not consider several important factors such as how well neighbors are linked among each other or how large the peak of interest is on the originally shocked page. This issues can only be adressed in a more complicated setup, such as my model in Section 5. However, the results from these descriptive regressions are based on minimal assumptions and provide guidance as to whether attention spillovers exist at all. They also highlight how far they carry over, and whether the shocks increased production. Finally, they allow me to provide a lower bound and an upper bound estimate of the aggregate spillover effects to be expected.

4 Data

This section surveys the data collection and the choice of control groups in Subsections 4.1 and 4.2. Subsection 4.3 describes the dataset I used and 4.4 shows first results. More details about the underlying database and extraction are in Appendix C.

4.1 Dataset and Definition of the Comparison Groups

The dataset is based on revision data from the entire German Wikipedia, which is provided as a so called "full-text dump." The history of the articles' hyperlink network was constructed by parsing the data to identify the links. A time-varying graph of the article network was built from the resulting tables. This graph provided the basis for sampling articles in my analysis. Furthermore information about page views and about the article content was added. Such additional information included, for example, the number of authors who contributed up to a particular point in time or the existence of a section with literature references. Since the data on page-views were collected after 2007, my data are based on 153 weeks of the the entire German Wikipedia's revision history between December 2007 and December 2010. The raw data are very large (terabytes) and hence it was not possible to conduct the data analysis using only in-memory processing. Instead, the data were stored in a relational database (disk-based) and queried using

[18] Note that these equations could also be written for indirect and further away neighbors (L_2, etc.)For L_2 (shown for large events in a robustness check)) the relevant coefficient would be $\phi_{2,0}^{L_2}$ for L_2.

Database Supported Haskell (DSH) (Giorgidze et al. (2010)).[19]

"Today's featured articles" were found by consulting the German Wikipedia's archive of articles that were advertised on Wikipedia's main page between December 2007 and December 2010. In German they are called "Seite des Tages."[20] To reduce the computational burden, I focus on articles that were advertised on the 5^{th}, 10^{th} or 20^{th} day of a month.[21] I identified all pages that received a direct link (L_1) from such a featured article more than a week before treatment. This ensures that my results are not driven by endogenous link formation.[22] For this set of pages, I use daily information on the contemporary state of the articles (page visits, number of revisions, page length etc.). I extracted the 14 days before and 14 days after the shock, which gives a total of 29 observations per page. For measuring attention spillovers I focus on the set of neighbors (L_1 or "one click away").[23]

4.2 Treatments and Choice of Control Group

To find a control observation for "Today's featured articles" I selected other featured articles, which were advertised later (e.g. "*next month's* featured article"). These articles fulfill the same requirements for the advertisement.[24] Furthermore they are equally eligible for treatment, as proven by the fact that they actually *were* treated, but at a different point in time. Finally, note that the focus of the estimation is on the *neighbors* of such articles. Neighbors are typically not themselves featured articles and are usually more similar to a randomly chosen article, than the treated featured article itself. Thus selecting neighbors of featured articles that were advertised at a different point in time, gives me the set $C_{1_{control}}$ which is similar in size, network structure and characteristics to the sampled pages (before the shock). The second control group is obtained by extracting the data based on treated pages a second time, but 42 days before the actual shock occurred. I refer to the articles in this "placebo-treatment" as $C_{1_{placebo}}$.

Table 7 (in appendix A) shows which featured articles were chosen by my procedure and included in the data. For reasons of space I show only 34 advertisements that correspond to the 10^{th} of a month. In general, the articles cover various topics such as innovations (e.g. the CCD-sensor), places (Helgoland), soccer clubs (Werder Bremen) or art history topics (Carolingian book illustrations). The first column of the table shows the number of neighboring articles that were linked to each featured article. The last three columns show the number of observations that received a link from an article before it

[19] This is a novel high-level language allowing the writing and efficient execution of queries on nested and ordered collections of data.

[20] http://de.wikipedia.org/wiki/Wikipedia:Hauptseite/Artikel_des_Tages/Chronologie_2008 etc.

[21] Three waves were extracted separately, to avoid the risk of temporal overlaps of different treatments.

[22] I thus only include pages that had a link before it was known that the start page will be shocked.

[23] I also analyzed articles 2 links away(L_2). However, the effects are small and not fully robust.

[24] Remember, that featured articles are longer, better linked, etc. than a randomly selected article.

was featured, separated by whether or not they belong to a time-series with actually treated observations.[25] The numbers range from 2,088 to 33,872.

4.3 Descriptive Statistics

The full data contain 908,628 observations from 15,732 pages[26] on the main variables. Table 1 shows the means of the main variables and their first differences by treatment status and before the treatment. The unit of observations is the page i on day t. Column 1 shows treated pages, Column 2 the control group and Column 3 shows the treated in the "placebo" condition, i.e. during a period without treatments. The time normalized variable is always negative, reflecting that the table focuses on the observations before treatment, reducing the number of observations to 337,974 (from 15,732 articles).

The mean page length is 6,736 bytes and 92 revisions. All three groups of articles have very similar characteristics. Even the first differences, i.e. the average changes in page length, number of links, number of authors, references and links to further info do not vary systematically between the groups. Moreover, the summary statistics of the first differences (variables starting with "Delta:") reveal that on a typical day nothing happens on a given page on Wikipedia. Less than 1 in 20 pages are edited on an average day. If anything, the "treated sample" is more likely to receive additional photos, but this affects less than 1 in 400 articles. The low levels of average activity highlight the necessity to use exogenous shocks as a focal lens for analyzing activity on Wikipedia, which shall be confirmed by the visual inspection of the direct and indirect effect of treatments. [27]

A summary of the full dataset and the distributions of the variables are provided in Table 6 in the data appendix. The table also shows median and selected percentiles of the distribution of the variables. Finally, Tables 8 and 9 show the summary statistics of the main "flow variables" (clicks and changes in the text ("first differences")) for treated and the comparison groups seperately. This further verifies that both groups had similar trends before the onset of treatment. As in Table 1, the neighbors of "Today's featured articles" show a somewhat higher editing activity already before their neighbor is advertised. However, these differences are insignificant and clicks are almost equal.

Figure 2 plots the aggregate dynamics around the day when the start node was shown on Wikipedia's main page. I plot the average clicks for the treated pages and direct neighbors. Each of the two figures contains three lines. The bold blue line represents

[25]Note, that each page shows up 29 times in the raw data and was sampled twice (placebo and real treatment), so that the number of corresponding pages (treatment or control) can be inferred by dividing the number of observations by 58.

[26]Since pages were observed also in the placebo condition, each page is sampled twice, and hence I observe 31,332 distinct time series.

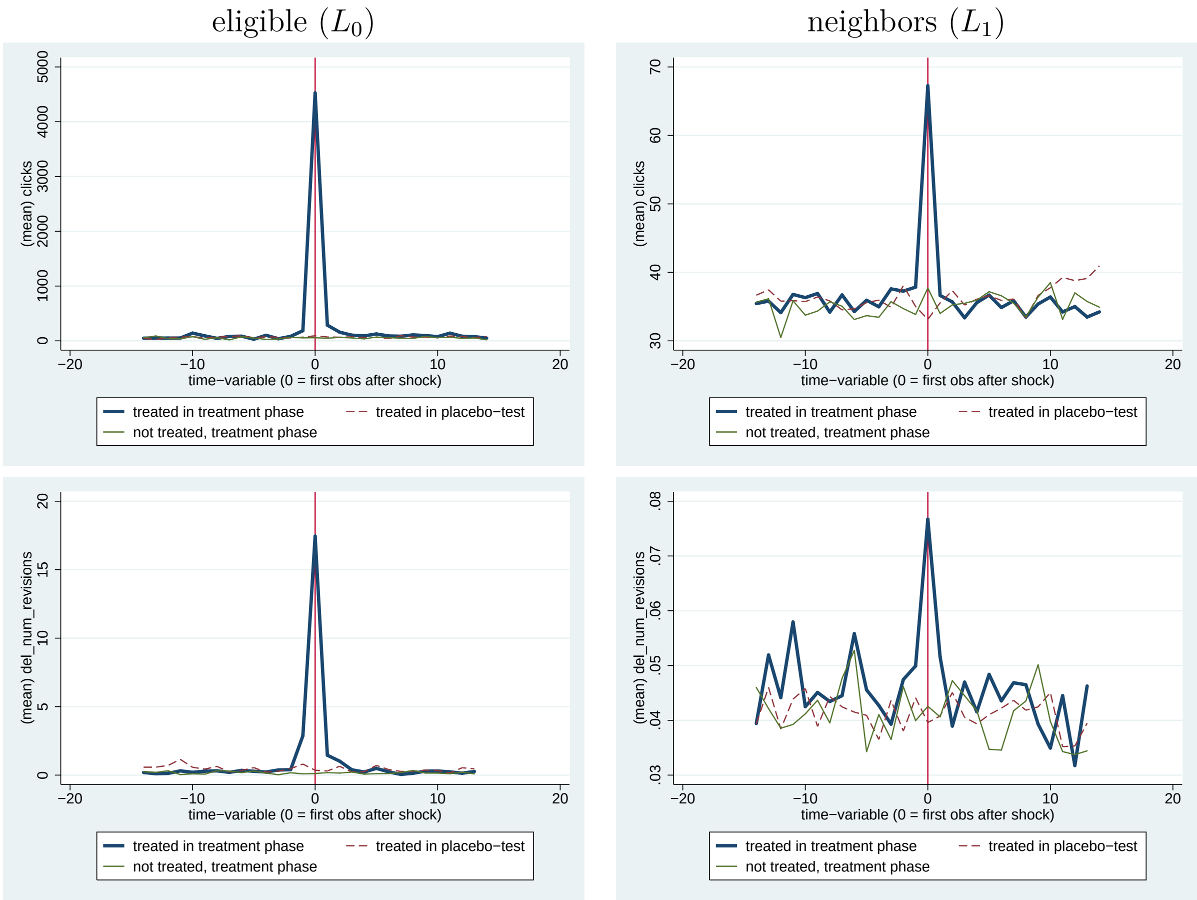
[27]Further descriptive analyses that compare treated and control groups before and during treatment show that the groups are very similar in their activity levels before the shocks occurred and that the control group did not change its behavior during treatment. These tables and their description were omitted for reasons of brevity. They are available in the online appendix.

Table 1: Summary Statistics on Articles by Treatment Status, before Treatment.

	1	2	3	Total
Clicks	36.01	34.39	35.89	35.48
	(133.6)	(146.9)	(130.0)	(136.5)
Length of page (in bytes)	6873.3	6523.3	6781.2	6736.0
	(6930.6)	(6696.2)	(6821.9)	(6824.4)
Number of images	1.805	1.947	1.772	1.836
	(6.126)	(5.923)	(5.814)	(5.958)
Dummy: literature section	0.305	0.295	0.299	0.300
	(0.461)	(0.456)	(0.458)	(0.458)
References (footnotes)	1.395	1.149	1.325	1.297
	(4.718)	(3.198)	(4.739)	(4.327)
External Links (web)	2.326	2.394	2.296	2.336
	(4.149)	(5.061)	(4.163)	(4.448)
Number of authors	32.41	30.78	31.79	31.70
	(34.76)	(34.26)	(34.00)	(34.35)
Number of revisions	94.05	89.89	92.25	92.17
	(131.7)	(133.8)	(129.1)	(131.5)
Number language links	15.16	13.54	15.26	14.71
	(19.08)	(18.90)	(19.11)	(19.05)
Links from Wikipedia	114.5	121.5	111.9	115.7
	(275.9)	(360.5)	(269.7)	(301.9)
Delta: Number of Revisions	0.0464	0.0421	0.0417	0.0435
	(0.420)	(0.391)	(0.357)	(0.390)
Delta: Length of page	2.464	2.209	1.957	2.211
	(166.6)	(116.9)	(166.4)	(153.3)
Delta: Number of authors	0.0160	0.0141	0.0144	0.0149
	(0.139)	(0.126)	(0.126)	(0.131)
Delta: Links from Wikipedia	0.0574	0.0644	0.0571	0.0594
	(0.871)	(0.571)	(0.603)	(0.701)
Delta: Number of images	0.00227	0.000699	0.000552	0.00120
	(0.402)	(0.0854)	(0.0507)	(0.245)
Delta: References	0.00189	0.00119	0.00144	0.00152
	(0.166)	(0.0895)	(0.132)	(0.135)
Delta: External Links (web)	0.000860	0.00115	0.000535	0.000834
	(0.124)	(0.0970)	(0.0829)	(0.103)

NOTES: The table shows the mean coefficients (sd in parentheses) of the main variables by treatment status. The unit of observations is the page i on day t. Column 1 shows treated pages and Column 2 the control group. Column 3 shows the treated in the 'placebo' condition, i.e. during a period without treatments. Column 4 show the averages for the entire dataset. The time variable is normalized and runs from -14 to 14.; no. of obs. = 337974; no. of start pages = 174; no. of articles = 15732.

Figure 2: Contrasting means of clicks over time: comparing 3 groups in one plot.



NOTES: The figure shows the results for featured articles that were advertised for a full day on Wikipedia's main page. The left column shows the average outcome on the directly treated pages (set "L_0" containing 187 pages total), the lower row for the pages one click away (set "L_1," which contains 15,732 pages). The upper row shows the average number of clicks the lower row shows the average number of edits.

the treated group or its neighbors when they were actually treated, hence "treated in treatment phase." The dashed red line represents the same group but during the placebo treatment at an earlier point in time. The thin green line ("not treated, treatment phase") shows the control group at the time when the real shock occurred .[28] The effect of the treatment is very brief and pronounced. Attention rises from typical levels, between 50 and 100 views, to approximately 4,500 views on average. It immediately returns to the old levels the day after treatment is administered. A very similar pattern can be observed for the neighbors where attention is almost twice as high as on a usual day and then falls back to the old levels, and this behavior is virtually mirrored by the number of revisions (edits). An important question is whether the network formation *process* is affected by the treatment. If the treatment drastically increased the number of links, I could not

[28]For greater ease of representation I included a graphical representation of only two variables. Also the other variables (cf. summary statistics) for these groups before and after treatment are available as tables upon request.

distinguish spillovers via existing links from additional spillovers through new links (cf. Comola and Prina (2013)). This is not a very big issue for "today's featured articles." Link formation at the treated articles increases by 0.2 new links on the advertised article, and at the neighbors by 0.04 new in-links (over 117 average in-links per article) on the day after treatment. It occurs a day after the peak in clicks and moves in parallel with a delayed peak in edits. I conclude that there is a small source of potential bias resulting from the edit activity, but that it is unlikely to affect viewership.

4.4 Difference-in-Differences

In what follows I present my estimation results and discuss their interpretation. The point of departure for the estimations (Section 3.4) for "Today's featured articles" is Equation 6. Recall that treatment takes place entirely inside Wikipedia[29] and is "completely local" since no two featured articles are advertised simultaneously. Only the advertised articles is directly affected and the large majority of these articles is about little known and very specialized topics. Hence, the variation in the neighbors is almost certainly a result of the processes inside Wikipedia, most importantly the presence of the link from the advertised page to the neighbor.

Table 2 shows the results for clicks in Columns 1-3 and the results for the number of added revisions in Columns 4-6. All the specifications are OLS panel regressions with page fixed effect and clustered standard errors (173 clusters on the event level). The table shows the coefficients for the period of the shock and the two subsequent days individually. Periods before the shock are represented in dummies that average over several days, and the periods later than two days after the shock are represented analogously. The reference group are days -14 to -8 before the advertisement. Until the onset of the event (days -3 to -1), we would expect insignificant effects for the shown cross terms. After the event has occurred a positive effect for treated nodes would be expected and for the neighbors such an effect would imply some form of spillover. Column (1) and (4) shows the results of the difference-in-difference estimations, which contrast the treated pages against the control group. Columns (2) and (5) show the contrast with a second comparison group, which is based on the treated articles themselves, but simulating a (placebo) treatment 42 days (i.e. 7 weeks) before the real shock. Column (3) and (6) shows the results of a simple before and after. This estimation does not depend on the validity of any comparison group and is, instead, valid if the shock is more important than underlying dynamics. The before and after estimates are almost equal and confirm that my findings are highly robust. The results do not depend on the choice of my comparison group.

I find a strong effect of the neighbor's advertisement on both average clicks and average edits. The size of the effect is estimated to be 28.5 to 34.6 additional clicks on the average

[29] Advertisement of featured articles on German Wikipedia is typically unrelated with media coverage.

Table 2: Relationship of clicks/added revisions and time dummies for direct neighbors of shocked articles in the 'featured articles' condition.

	clicks			Δ revisions		
	(1) compare control	(2) comp. placebo	(3) before after	(4) comp. control	(5) comp. placebo	(6) before after
Before: days - 7 to -4	0.821	0.486	-0.181	-0.000	0.002	0.001
	(1.342)	(1.222)	(1.066)	(0.004)	(0.004)	(0.003)
Before: days - 3 to -1	1.811	2.026	1.910	0.000	-0.001	-0.001
	(2.408)	(1.990)	(1.462)	(0.005)	(0.004)	(0.003)
t = 0	28.546***	34.577***	31.603***	0.030***	0.033***	0.030***
	(5.948)	(5.731)	(5.573)	(0.009)	(0.008)	(0.007)
t = 1	1.632	1.535	0.974	0.007	0.006	0.005
	(2.146)	(2.318)	(1.565)	(0.007)	(0.007)	(0.005)
t = 2	-0.569	-1.189	0.028	-0.013*	-0.011	-0.007*
	(2.768)	(2.395)	(1.910)	(0.007)	(0.007)	(0.004)
After: days 3 to 6	-2.170	-0.376	-0.531	0.002	-0.000	-0.001
	(2.052)	(2.296)	(1.359)	(0.004)	(0.004)	(0.003)
After: days 7 to 14	-0.639	0.207	0.207	0.001	-0.001	0.001
	(2.593)	(2.794)	(1.953)	(0.007)	(0.007)	(0.005)
Time Dummies	Yes	Yes	No	Yes	Yes	No
Mean dep. Variable	36.208	36.559	37.276	0.045	0.045	0.047
Observations	346104	371382	186384	346104	371382	186384
Number of Pages	15732	16881	8472	15732	16881	8472
Adj. R^2	0.002	0.003	0.004	0.000	0.000	0.000

NOTES: The table shows the results of the reduced form regressions estimating the ITE . Columns (1)-(3) show the results for clicks and Columns (4-6) for new edits to the articles. (1) and (4) contrast treated and comparison group; Columns (2) and (5) show the comparison of treated articles with themselves but seven weeks earlier (placebo treatment). Specification (3) and (6) show a simple 'before and after.' Fixed Effects Panel-Regressions with heteroscedasticity robust standard errors. The unit of observations is the outcome of a page i on day t. The time variable is normalized and runs from -14 to 14.; Only crossterms on and shortly after the day of treatment are shown individually, all others are shown in groups. Reference group t-14 to t-8; standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; no. of obs. = 689,304; no. of clusters = 174; no. of articles = 15,732.

neighbor page on the day of treatment. In Columns 4-6, I estimate between 0.029 and 0.032 additional revisions one day after the treatment of the neighbor page. This is an important effect and it is worth noting two things here: First, the effect is very small in absolute terms and corresponds to one additional edit per thirty pages. Second however, this is an increase in contribution activity of eighty to one hundred per cent. Also the number of authors experiences a spike paralleling the one for edits. This is shown in Column 5 of Table 3 (Subsection 5.2) and documents that also new contributors edit the articles. Usually less than 1 in 50 articles (on average) is edited by a new author (who never edited the article before). During treatment 1 in 30 of the neighbors are edited by a new author (a 72% increase).

Note that I took several steps to avoid confounding factors that may threaten the validity of my results. I considered only links that had been in place *a week before the treatment* to avoid potentially endogenous link formation during treatment. When a page is found to lie in both the treatment and control groups it is excluded from the estimation, because including such pages will bias the estimated coefficients towards zero. Extremely broad pages with a very large number of links (e.g. pages that correspond to years) were excluded from estimation to avoid biases from oversampling. Finally, I use the seven observations from two weeks before treatment (days -14 through to -8) as the reference group in the estimations and I include only flow variables such as views, new revisions, new authors etc. to guarantee that my results are not driven by any anticipation effects.[30]

I further test the robustness of my difference-in-differences results. First, I excluded the advertisements of "featured articles" that took place on a weekend. If a part of my effect were driven by user-idleness, the spillovers should be smaller during the working days. Results, which are shown in Table 19 are virtually identical. Next, I check whether my sampling on 3 days of the month matters, by further restricting the sample to advertisements that came from the 10^{th} day of each month (cf. Table 20).[31] This reduces the number of advertisements to 34, but otherwise reveals the same patterns as Table 18, just at lower significance levels. In both robustness checks the number of authors moves largely in parallel with the number of revisions, indicating that twice as many new authors as usual edit the article due to the treatment of its neighbor. On the one hand this is a large effect in relative terms, on the other hand it means that only one in seventy articles receives an additional edit by a new author. More robustness checks included regressing against all comparison groups simultaneously and using different samples or resolutions. They do not convey additional insight and are available in the online appendix.[32]

In Appendix D.2, I present results for a different dataset from natural disasters as a robustness check. This addresses the concern that anticipation effects cannot be entirely ruled out for "Today's featured articles." Since it is arguably not possible to anticipate disasters, I extracted and estimated the spillovers in attention for disasters. Anticipation is arguably not possible for disasters and find very similar coefficients. The extraction of the data and the results are discussed in Appendix D.2.

5 Attention Spillover

Beyond measuring the size of the indirect treatment effect (ITE), I want to quantify the size of the structural ("day to day") spillovers of views between Wikipedia articles. In this section, I augment the well known linear-in-means model for peer effects in networks

[30]Anticipation effects cannot be entirely ruled out in the "Today's featured article" condition. Sophisticated users can obtain information about future advertisements. In fact, the main editors of the daily advertisement of articles, increase their activity on the articles about a week before treatment. This is necessary to fit them into the corresponding box on Wikipedia's main page. After carefully studying this process, I am not very concerned about this feature of the data, because, if relevant, increased activity pre treatment generates a downward bias of the difference-in-differences. Moreover, the magnitude of the day-0 effect suggests that the influx of attention is due to readers who did not anticipate the advertisement. Nevertheless, in Appendix D.2, I present results for a different dataset from natural disasters as a robustness check.

[31]These were also the main results of a previously circulated version.

[32]https://sites.google.com/site/kummersworkingpapers/spilloversonlineappendix

(Manski (1993), Bramoullé et al. (2009)). I incorporate local exogenous shocks into the model to show how they can help identifying spillovers (or peer effects). Here, I provide the point of departure and the main results. Details and derivations can be found in Appendix E.[33]

5.1 Identifying the Spillover Parameter

I am interested in measuring to which extent links channel attention from articles to their neighbors.[34] This can be modeled using the established empirical "linear-in-means" approach of the type discussed in Manski (1993). The network version of this model (cf. below) can be derived from an adapted version of the "random surfer model." This model of individual surfing behavior lies beneath Google's Page Rank measure (Page et al. (1999)). This measure can quantify the importance of pages in networks and provides a natural starting point. I show how to connect the models in an online appendix.[35]

Manski shows that the coefficient of interest is generally very hard to identify and the empirical model is my point of departure:[36]

$$y_{it} = \alpha \frac{\sum_{j \in P_{it}} y_{jt}}{N_{P_{it}}} + X_{it-1}\beta + \gamma \frac{\sum_{j \in P_{it}} X_{jt-1}}{N_{P_{it}}} + \epsilon_{it}$$

Here y_{it} denotes page views in period t, and X_{it-1} are i's observed characteristics at the end of period $t-1$.[37] Let P_{it} be the set of i's peers in period t and $N_{P_{it}}$ its cardinality. The parameter of interest is α: It captures the effect of the performance of i's peers. In the present context it measures how an article's views depend on the average views of adjacent articles, and it is assumed to be homogenous across articles. The other coefficients are the vectors β and γ. Generally, β accounts for i's own characteristics and γ measures how the peers' average characteristics affect i's performance. In this paper β accounts for how the page's own length or quality might affect how often it is viewed. γ captures how length and quality of neighboring pages affect views of page i. Bramoullé et al. (2009) suggest a more succinct representation based on vector and matrix notation:

$$\mathbf{y_t} = \alpha \mathbf{G} \mathbf{y_t} + \beta \mathbf{X_{t-1}} + \gamma \mathbf{G} \mathbf{X_{t-1}} + \epsilon_\mathbf{t} \quad where \quad \mathbf{E}[\epsilon_\mathbf{t} | \mathbf{X_{t-1}}] = 0$$

Now $\mathbf{y_t}$ is n x 1, and $\mathbf{X_t}$ is n x k. $\mathbf{G}$ is the n x n row normalized adjacency matrix. For its elements I maintain the standard notation, such that $G_{ij} = \frac{1}{N_{P_i}-1}$ if i receives a

[33]The derivations involve quite heavy notation, but are otherwise relatively straightforward.

[34]The mechanism we have in mind, is that attention from article A can be diverted to article B if a link exists. This is interesting, since some of the users who get to see B might later start to edit it.

[35]Available upon request. I do not present this model here. It uses additional notation and is not the contribution of this paper.

[36]Note that it is easy to add a fixed effect to the model, but that it will be eliminated when taking differences. Consequently, I omit it for ease of notation.

[37]Note, that I can observe the current state of a Wikipedia article once a day at a fixed time.

link from j and $G_{ij} = 0$ otherwise.

I augment this model to incorporate exogeneous variation, by including a vector of treatment. For simplicity, I assume this vector to take the value of 1 for treated nodes and the value of 0 otherwise.

$$(7) \qquad \mathbf{y_t} = \alpha \mathbf{G} \mathbf{y_t} + \mathbf{X_{t-1}} \beta + \gamma \mathbf{G} \mathbf{X_{t-1}} + \delta_1 \mathbf{D_t} + \epsilon_\mathbf{t} \quad \textit{where} \quad \mathbf{E}[\epsilon_\mathbf{t} | \mathbf{D_t}] = 0$$

For the treated side $\mathbf{D_t}$ is an n x 1 vector consisting of zeros and ones that indicates which nodes are treated. On the untreated subnetwork we have $\mathbf{D_t} = \mathbf{0}$, a vector of zeros (a "mini population treatment"). Formally this is written as $\mathbf{D_t} = \mathbf{e}_{\ell \mathbf{0}}$; that is, a vector of zeros and a unique one in the coordinate that corresponds to the treated node.

I would like to stress that this result does not require an exogenous formation process of the network $\mathbf{G}$ ($\mathbf{E}[\epsilon_\mathbf{t} | \mathbf{X_t}] = 0$). Rather, the network is assumed to be stable in the short run. More importantly, it must be exogenous which of the eligible nodes gets treated *today* ($\mathbf{E}[\epsilon_\mathbf{t} | \mathbf{D_t}] = 0$).[38] This an entirely different source of identification than Bramoullé et al. (2009). Moreover, there will be no requirements needed concerning the linear independence of $\mathbf{G}$ and $\mathbf{G^2}$.

The reduced form expectation conditional on "treatment" is given by:

$$(8) \qquad \mathbf{E}[\mathbf{y_t} | \mathbf{D_t}] \quad = \quad (\mathbf{I} - \alpha \mathbf{G})^{-1} [(\beta + \gamma \mathbf{G}) \mathbf{E}[\mathbf{X_{t-1}} | \mathbf{D_t}] + \delta_1 \mathbf{D_t}]$$

The matrix $\mathbf{I}$ is the n x n identity matrix, and $(\mathbf{I} - \alpha \mathbf{G})$ is invertible if α is small enough to prevent the system from exploding.[39] Define the set of observations in the subnetwork where treatment occurs in t by the subscript ℓ and consider a comparison group in which no node is treated (denoted by c). If these sets of nodes can be observed one period earlier, a difference-in-differences can be computed. Rewriting the differences-in-differences in terms of the reduced form from above gives:

$$\mathbf{DiD} := \{\mathbf{E}[\mathbf{y}_{\ell,\mathbf{t}} | \mathbf{D}_{\ell,\mathbf{t}}] - \mathbf{E}[\mathbf{y}_{\ell,\mathbf{t-1}} | \mathbf{D}_{\ell,\mathbf{t-1}}]\} - \{\mathbf{E}[\mathbf{y_{c,t}} | \mathbf{D_{c,t}}] - \mathbf{E}[\mathbf{y_{c,t-1}} | \mathbf{D_{c,t-1}}]\}$$

In what follows I consider the placebo treatment (-S periods) for comparison ($y_{c,t} = y_{\ell,t-S}$) and assume that:

1. Network $\mathbf{G}$ is stable between $t - S$ and t, i.e.: $\mathbf{G}_{\ell,\mathbf{t}} = \mathbf{G}_{\ell,\mathbf{t-1}} = \mathbf{G}$ and $\mathbf{G_{c,t}} = \mathbf{G}_{\ell,\mathbf{t-S}} = \mathbf{G}$.

[38] In the present application, all "eligible" nodes (the featured articles in my treated and control groups) are assumed to be equally likely to be advertised. They are the nodes in the group L_0. Neighbors (in L_1) are typically not featured. Hence they are not eligible and naturally less likely to be themselves treated.

[39] Precisely, for invertibility is ensured if $\alpha < 1$ (Bramoullé et al. (2009)) and the infinite sum is well defined if α is smaller than the norm of the inverse of the largest eigenvalue of $\mathbf{G}$ (Ballester et al. (2006)). For Wikipedia these assumptions are obviously satisfied.

2. In expectation, the period on period changes of the pages between $t-1$ and t are the same as from $t-S-1$ to $t-S$ (cf. Appendix for a precise formalization).

3. There exists a local exogenous shock ($\mathbf{E}[\epsilon_\mathbf{t}|\mathbf{D_t}] = \mathbf{0}$)

4. The Stable Unit Treatment Value Assumption (SUTVA) holds on the level of subnetworks: The non-treated subnetwork is unaffected by treatment.

5. Treatment does not affect the independent characteristics X of the treated node.[40]

Result 1: Suppose that assumptions 1-5 are satisfied and α is small enough for $(\mathbf{I}-\alpha\mathbf{G})^{-1}$ to be well defined. Then the observed difference-in-differences estimates are a function of the shock and repeated rounds of the spillover, i.e.:

$$\mathbf{DiD'} \quad = \quad \delta_1 \mathbf{D'_t}(\mathbf{I} + \alpha\mathbf{G} + \alpha^\mathbf{2}\mathbf{G^2} + \alpha^\mathbf{3}\mathbf{G^3} + ...)'$$

Proof: For a proof of Result 1 and a more detailed discussion of the assumptions, please refer to Appendix E.3.

In words, this result means that the node is affected by both treatment and second and higher order spillovers, the positive feedback loop that ensues as the neighbors increase their performance in sync with their peers. Instances of higher order effects are $\alpha^2\delta_1$ in the second round or $\alpha^3\delta_1$ in the third round and so on.[41] The other important factor is whether and how often spillovers of a given order q arrive. This depends on the number of indirect paths of length q that go from the shocked node $\ell 0$ to any focal node j.[42]

Note the close relationship to the Page Rank measure (Page et al. (1999)) and the Katz-Bonacich centrality (Ballester et al. (2006), Calvó-Armengol et al. (2009)). Ballester et al. (2006) aim at identifying the "key player" of a network. Like in their framework, the *number of channels* for indirect spillovers matters. However, for measuring spillovers in a "mini population treatment" we care about the reverse direction, the quantity that spillovers from the shocked node to any other node.

My result shows that the difference-in-differences approach alone will not directly reveal α, the social parameter of interest, because nodes might have a feedback effect on each other. The neighbor's change in performance (due to the original impulse) will affect the neighbors' neighbors, but also feed back to the originally treated ℓ_0-node. The estimator will observe all the changes in outcome at the end of this process, when all higher order spills have taken place.

[40] The effect should at least be negligible, relative to the effect on the outcome. Else, identification of the spillover is threatened. X may adjust over time, though.

[41] Note that I am considering the homogeneous network, so all spillovers have the same magnitude.

[42] In the proof I need to assume that the network formation *process* is not affected by the treatment. I checked this assumption in my "today's featured article application" and verified, that link formation remains on low levels. If anything, there is an increase by 0.2 in-links per article which parallels the peak in edits, but not with clicks. I conclude that this is an acceptably small source of potential bias.

While this treatment effect is the object of interest in many applications, I study the spillover parameter α. Generally, measuring α requires knowing the complete link structure and is thus not necessarily feasible. However, in my setup, limited information about the link structure suffices to acquire additional information about the parameters. In the following two subsections I first discuss how to estimate the spillover parameter (or more generally the social parameter) if the network is known. After that, I derive upper and lower bound estimates for the parameter if no information about the network is available.

5.1.1 Estimating the Spillover Parameter: Known Network Structure

If the network structure can be observed, the peer effect parameter α can be backed out by computing the higher orders of the network graph (**G**-matrix). To know how many spillovers arrive in each round, it suffices to focus on the entries $\mathbf{G}_{ij}, \mathbf{G^2}_{ij}, \mathbf{G^3}_{ij}, etc.(i = \ell_0)$ that document the number of paths via 1, 2, 3 and more links from the treated node to the neighboring node in question. With this information it is straightforward to compute by how much the observed effect at the node in question has to be discounted and to use this information to compute the true average effect.

5.1.2 Bounds of the Spillover Parameter: Unobserved Network Structure

If the network structure is unobserved, it is possible to obtain bounds of the spillover parameter α. I now briefly sketch how to obtain these bounds. They merely require separate estimation of the direct treatment effect (L_0 vs. C_0) and the indirect treatment effect (L_1 vs. C_1). In many empirical settings, researchers cannot observe both randomization and the network *together*, but only either of them.[43] My bounds are relevant in such empirical settings.

The key idea is selecting two specific "extreme" types of network as benchmarks, as shown in Figure 3. A well chosen benchmark network either minimizes or maximizes second and higher order spillovers. For the upper bound estimate, I use a directed "star network", which only links "outward" from ℓ_0 to $\ell_1 \in L1$.[44] For the lower bound estimate I use a fully connected network, where higher order spillovers are maximized. More details and the proofs are provided in Appendix E.4.

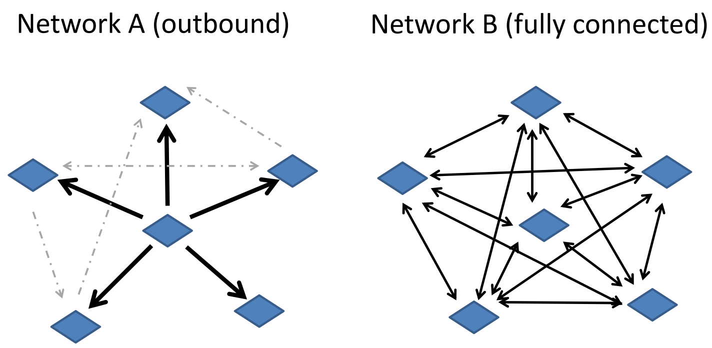
Upper Bound: Ignoring higher order spillovers,[45] we can obtain an upper bound estimate for the direct treatment effect ($\overline{\delta_1}$) by applying the difference-in-differences esti-

[43]In contrast, a separate comparison of eligible and non-eligible nodes in randomly treated communities or networks (without network information) can frequently be observed.

[44]For this benchmark we ignore any existing links among L_1 nodes.

[45]Alternatively, maintain the assumption that we can observe the nodes' performance before any higher order spillovers arrive at the treated node

Figure 3: Schematic representation of the two extreme networks, used to compute the upper and lower bound estimates of the parameters of interest.



NOTES: The " outbound network" (left) is used to obtain the upper bound estimate. It is a directed network with only "outward bound" links. This implies ignoring any existing links among L_1 nodes. Holding the number of nodes and the observed ITEs fixed, the social parameter will be estimated to be largest in this type of network. The fully connected network (right), is the benchmark case from which the lower bound of the social parameter can be estimated.

mator on the level of directly treated nodes (L_0) and a suitable comparison group (C_0). After that I can estimate the upper bound for the parameters for spillovers ($\overline{\alpha}$) by combining it with a second DiD estimator at the neighbor level. Let $DiD_{(\ell_a - c_a)}$ denote such a point estimator, ($a \in \{0,1\}$), of the respective average effects of the shock on the start nodes (L_0 or C_0) or the direct neighbors (L_1 vs. C_1):

$$\begin{aligned}\overline{\delta_1} &= \widehat{DiD_0} = \widehat{\Delta\ell_0} - \widehat{\Delta c_0} \\ \overline{\alpha} &= \frac{\widehat{DiD_1}}{\widehat{DiD_0}} N_{P_{\ell_1}}\end{aligned} \tag{9}$$

- $\widehat{\Delta\ell_0} := \frac{1}{N_{P_{\ell_0}}} * \sum_i (y_{i,\ell_0,t=1} - y_{i,\ell_0,t=0})$
- $\widehat{\Delta c_0} := \frac{1}{N_{P_{c_0}}} * \sum_i (y_{i,c_0,t=1} - y_{i,c_0,t=0})$

with $\widehat{DiD_1}$ and $\widehat{DiD_0}$ being the 1 x 1 estimator of the average effect of the shock. In analogy to $\widehat{DiD_0}$, $\widehat{DiD_1}$ is given by $\widehat{\Delta\ell_1} - \widehat{\Delta c_1}$, and the definitions of $\widehat{\Delta\ell_1}$ and $\widehat{\Delta c_1}$ correspond to those of $\widehat{\Delta\ell_0}$ and $\widehat{\Delta c_0}$. In the estimations of the previous section, the $\widehat{DiD_1}$ is estimated by $\widehat{\phi^{L_1}_{2,0}}$ from Equation 6 and $\widehat{DiD_0}$ by $\widehat{\phi^{L_0}_{2,0}}$ (Equation 5). This upper bound estimator would be suitable under the potentially quite strong assumption that higher order spillovers are negligible. I proceed to show how to compute the lower bound estimates under the assumption of *maximal* second order spillovers. The lower bound gives an idea of the maximal size of the problem that might result from trusting the easily computed upper bound estimates.

Lower Bound: It is also possible to compute a lower bound estimate for α and δ_1. This bound can be obtained by imagining that the network is a single connected component, i.e. every node links to every other node, assuming that all effects are of the same sign, strictly ordered and positive.[46] Further computations in Appendix E show that in a network with n nodes, the lower bound of the estimator for α is characterized by the solution to the following quadratic equation:

$$\underline{\alpha}^2 - [\frac{\widehat{DiD_0}}{\widehat{DiD_1}} + (n-1)]\underline{\alpha} + (n-1) = 0 \tag{10}$$

This equation has two solutions, one of which lies between 0 and 1. The closed form solution for $\underline{\alpha}$ is hence given by:

$$\underline{\alpha} = \frac{1}{2}[\frac{\widehat{DiD_0}}{\widehat{DiD_1}} + (n-1)] - \sqrt{\frac{1}{4}[\frac{\widehat{DiD_0}}{\widehat{DiD_1}} + (n-1)]^2 - (n-1)} \tag{11}$$

Recall that all the quantities required are readily available from the reduced form estimations. $\widehat{DiD_1}$ corresponds to $\widehat{\phi_{2,0}^{L_1}}$ and $\widehat{DiD_0}$ is estimated by $\widehat{\phi_{2,0}^{L_0}}$. In Appendix E.4 I provide the proofs and explain how this bound is derived. Which of the estimates is more accurate will depend on the size of the spillover effect, but to a very large extent also on the real network structure and the number of nodes.

A closer examination of Result 1 reveals that the upper bound estimator is suitable if the researcher has reasons to make the (potentially strong) assumption that higher order spillovers are negligible. It would also be appropriate in networks with very sparse connections among its members. The lower bound estimator might be more suitable if the researcher believes the network to be highly connected and expects the spillover coefficient to be relatively large.[47] The bounds have several limitations (cf. Appendix E.4) and for some applications the bounds might turn out to be too wide to be informative. Still, taken together, the bounds can provide a useful first characterization of the spillover parameters in question.

5.2 Estimation of the Spillover Parameter

To estimate the spillover effect, I combine the difference-in-differences estimates from Section 4 with the concepts from the previous subsection (5.1.2). This allows me to obtain the structural spillover parameter (α in the model), which estimates how much an

[46] Positivity is without loss of generality. The precise assumption is $DiD_0 > DiD_1 > HO^B > 0$, as stated and explained in Lemma 1

[47] So large that α^2 and α^3 are still sizeable.

article's attention depends on the average attention of its neighbors. Once the difference-in-differences is computed, the additional computations are easily implemented.

Table 3: Relationship of clicks/added revisions and time dummies.

	epicenter (L0)		direct neighbors (L1)		
	(1) clicks	(2) clicks	(3) clicks	(4) new edits	(5) new authors
Before: days - 7 to -4	7.053	-1.198	0.486	0.002	-0.000
	(12.579)	(12.614)	(1.222)	(0.004)	(0.001)
Before: days - 3 to -1	33.263	32.042	2.026	-0.001	-0.002
	(26.940)	(27.647)	(1.990)	(0.004)	(0.002)
t = 0	4451.648***	4425.120***	34.577***	0.033***	0.015***
	(1000.123)	(1000.523)	(5.731)	(0.008)	(0.003)
t = 1	219.809***	207.400***	1.535	0.006	0.001
	(62.211)	(66.278)	(2.318)	(0.007)	(0.002)
t = 2	77.166*	84.088**	-1.189	-0.011	-0.004**
	(41.330)	(38.779)	(2.395)	(0.007)	(0.002)
After: days 3 to 6	29.716*	34.317	-0.376	-0.000	-0.001
	(17.963)	(21.008)	(2.296)	(0.004)	(0.001)
After: days 7 to 14	17.749	-17.890	0.207	-0.001	-0.001
	(36.823)	(37.126)	(2.794)	(0.007)	(0.002)
Time Dummies	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	173.800	181.277	36.559	0.045	0.015
Observations	4114	4092	371382	371382	371382
Number of Pages	187	186	16881	16881	16881
Adj. R^2	0.169	0.169	0.003	0.000	0.000

NOTES: The table summarizes the results of the reduced form regressions estimating the spillovers in clicks and edits. Columns (1)-(2) show the results for the direct effect of treatment on clicks (ATE) and Columns (3) the spillover to direct neighbors of the articles (ITE). and Columns (4) show the conversion of the spillover in clicks to new edits at the direct neighbors and Column (5) shows the number of new author's that contributed to the articles (at the neighbors). All Specifications compare the nodes in treatment to the 'placebo treatment' 7 weeks earlier. Fixed Effects Panel-Regressions with heteroscedasticity robust standard errors. The unit of observations is the outcome of a page i on day t. The time variable is normalized and runs from -14 to 14.; Only crossterms on and shortly after the day of treatment are shown individually, all others are shown in groups. Reference group $t-14$ to $t-8$; standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1;

Table 3 summarizes the results of the difference-in-differences estimations in view of the attention spillover and the ensuing content production. The first two columns show the results for views of the treated article in L_0. The coefficient for $t = 0$ corresponds to $\widehat{DiD_0}$ as defined by $\widehat{\phi^{L_0}_{2,0}}$ in Equation 5. In Columns 3 the dependent variables are views of the neighbors. We are again interested in the coefficient in period 0 which is $\widehat{DiD_1}$ ($\widehat{\phi^{L_1}_{2,0}}$ in Equation 6). Column 4 repeats the estimation from Column 3, but for edits on neighboring articles. It compares the treated and the placebo group, as in Table 2. In Column 5, I add results for the number of editors. As can be seen in the first two columns, the estimated direct effect of treatment $(\widehat{DiD_0})$ is approximately 4,450 views, depending on the comparison group, and $\widehat{DiD_1}$, the estimated difference in difference, is 34.5 views.

5.3 Bounds for the Spillover Parameter

Unfortunately I cannot present the precise structural estimator, since the full matrix $\mathbf{G}$ formed by the German Wikipedia is too large to compute in memory. Hence, I cannot solve for $\mathbf{G}^2$ and higher orders of the link matrix. However, it is possible to present upper and lower bound estimates of the structural parameter α that was discussed in Subsection 5.1 and derived formally in Appendix E.

I proceed to illustrate how to retrieve the bounds for the structural parameters in my application. The number of peers is evaluated at the median (34 for neighbors of "Today's featured articles"). The rest reduces to a back of the envelope calculation for the upper bound of the social/spillover parameter α and the effect of the shock δ_1. I use Equation 34: $\overline{\delta_1}$ is directly estimated to be 4,450. The estimate of $\overline{\alpha}$ is 0.241 based on "featured articles". Computing the lower bound estimates is not much more involved: Plug the estimates and the number of nodes into the closed form solution given in Equation 40. This gives the point estimator for the lower bound $\underline{\alpha} = 0.190$. To conclude this section I quantify the implications of my estimates: Increasing the average clicks on the neighbors by ten, would result in an increase of 1.9 to 2.41 clicks on the page. Hence, an increase aggregate viewership of the neighbors by 200 is predicted to result in 1.2 additional views on the target page.[48]

6 Content Generation

In this section I scrutinize where and how much additional content is generated due to the advertisements. I first discuss the aggregate effects of "Today's featured articles" treatment and illustrate them graphically. Second, I analyze which article and network characteristics mediate attention spillovers and how attention is converted into content.

6.1 Aggregation over Neighbors

First I analyze aggregate changes in clicks and revisions over all neighboring articles. Figures 4 and 5 show the average activity when summing over all neighbors (L_1) in the 93 analyzed "Today's featured article" clusters. On average I observe 3,000 clicks additional clicks on all neighbors taken together (Figure 4). Given that the average treated articles received an additional 4,500 clicks this corresponds to a 67 percent conversion of "first clicks" on the treated page to "second clicks" on one of the neighbors. Two thirds of the average visitors clicks on a link. Moreover, the total number of revisions on the neighboring pages (Figure 5) increases from 4.5 to 7.5 on the day of treatment. This

[48]The computation follows from equation 9 and from evaluating neighbors and clicks at the median. Consequently 200 aggregate views correspond to 5 more average views, because the median page has 34 neighbors. The quantification is based on $\overline{\alpha}$.

implies an aggregate conversion ratio of 1,000:1 additional clicks to edits on a neighbor. Where usually one in 22 articles is edited, it is one in 12.5 on the day of treatment. Interestingly, these findings are in line with the average "facebook-engagement rate", which is typically just below 0.01.[49]

6.2 Allocation of Edits and Heterogeneous Effects

In this subsection I analyze how article and network characteristics mediate the spillovers of attention and the associated content generation. In the first subchapter I analyze how article characteristics, such as length or connectedness, are related to content generation. In the second subchapter I analyze how the spillover is associated to factors that are determined on the shocked (L_0) articles. I focus on a link's position and the number of views to the shocked pages as mediating factors to the size of the spillover.

For the two analyses in this subchapter, I put together an augmented cross-sectional dataset that evaluated the attention peaks of the treated articles (L_0) and their neighbors (L_1) on the day of treatment. Moreover I parsed the text in the treated articles to add information on the links' order of appearance.[50] I also added the information about article properties and on how intensely an article was linked to other L_1 neighbors (closed triads). Finally, for each article I added information about the total amount of attention on the corresponding treated (L_0) article and how many links could be found in that text.[51]

Most interestingly, the analysis of the effect of link position on the click-throughs and edits suggests that attention and relatively short edits are influenced by the order in which links appear, but that deep edits are not.

A caveat of the subsequent crosssectional analysis lies in many unobserved factors, such as relevance, challenging topics, etc. While these are dealt with in the previous difference-in-difference, this is no longer possible in the cross-section. Hence, these regressions might introduce endogeneity. However, since these correlations are relevant, I report them with a warning that they cannot necessarily afford a causal interpretation.

6.2.1 Article Characteristics and Edits

I now report how article characteristics influence the spillovers, when accounting for differences in the length or connectedness of the neighbors. This not only sheds light on where content is provided after attention shocks, it also serves as a test of the assumption

[49] http://www.michaelleander.me/blog/facebook-engagement-rate-benchmark/. The benchmark measures, how many of a user's friends and followers react to their posts. The conversion is also in line with German Wikipedia's conversion rates in the yearly funding campaign.

[50] I evaluated the treated article's version that was valid at the beginning of treatment.

[51] I applied an outlier-correction(removing .5% extreme observations of each dependent variable). Moreover, a coding issue didn't allow to match the information in all articles with a German "Umlaut". However, the variables that I can compare are similar in both sets of observation.

that attention spillovers are homogeneous, as is assumed in my model. The results of this analysis are shown in the first three columns of Table 10. In these regressions I added variables that account for an article's length, how well it is generally linked and how closely they are linked to the shocked articles (by counting closed triads with other neighbors). Columns 4-6 show the results of an analysis that considers short articles ("stubs") seperately.

Splitting the sample into very long, very short and other articles, I do not find a significant difference between the long articles and the other groups. I get a positive but insignificant point estimate for page views. (Column 1). I also differentiate well-connected articles with many links and poorly connected ones with few and find a negative relationship for "well-linked" pages. On the other hand, there is a statistically significant positive relationship for nodes that show high clustering relative to the shocked node (neighbors, that get many links also from other neighbors). The point estimate indicates a higher click-through rate onto those articles, but note that these clicks might be indirect click-throughs (forwarded from neighbors of the shocked node). Edits are less likely for "well-linked" articles and for short articles. Length and high clustering are not associated with any significant increases in editing or page length.

An interesting pattern emerges, when I consider only "stubs", i.e. pages that do not exceed a length of 1500 bytes. I find that the increase is almost the same as on average, but that attention is less likely to convert into edits (Columns 4-6 of Table 10). Short articles are viewed 25 times more often than on an average day, this corresponds to approximately 80 percent of the average spillover. However, in absolute terms the increase in edits is smaller (.017 vs. .034, in separate estimation). On the other hand, stubs generally have a much lower probability to be viewed or edited, these increases are *larger in relative terms* than for normal articles.[52]

Table 4: Summary of results: Conversion of attention to content generation.

	avg.	length of target		link position	
	all articles	long articles	short articles	link on top	link in end
click-through	0.008	0.008	0.006	0.012	0.004
conversion to action	0.0013	0.0014	0.0005	0.0008	0.0013

NOTES: The table contrasts the conversion of clicks to content for different types of articles. Short articles have similar click-through rates, but a lower conversion of attention for content. Note that long articles have a slightly higher click through than average, the difference is lost in rounding. The groups are not equally weighted, hence the categories do not give the average. Finally, note that articles linked higher on a page get more activity, despite the lower conversion rates. Source: own calculations, based on tables presented in the paper and the online appendix.

Altogether my findings indicate that the attention spillovers do vary systematically by

[52]The probability of being viewed increases by 300% (vis a vis 100%) for the average neighbor. The chances for an edit increases by 500% vis a vis an 80% increase over baseline levels. This much stronger relative effect in the number of edits indicates that contributors do make a greater effort (in comparative terms) to contribute to pages where the existing content is limited.

the four mentioned properties. However, for "well-linked" and "highly-clustered" pages I find counterveiling effects, which suggest that articles about a very broad topic are less clicked unless closely related (as indicated by the positive association with clustering). This interpretation would be in line with my findings on editing.

6.2.2 Link Position and the Size Spillovers.

In this subsection, I analyze the role of a link's position in the outgoing treated articles. Links that appear early in the shocked (L_0) article are expected to channel more attention to the linked neighbor than links in the end. This expectation is formed for at least three reasons: First, users - reading from top to bottom - might click on a link or just stop reading before they discover a relevant link below. Second, featured Wikipedia articles have a summary of the contents upfront. After reading the summary, users might decide to click on a link from the summary. They might like the current article less than another link in the summary. Third and finally, highly relevant links are more likely to appear early (e.g. because the summary is on top). While the first two reasons would be in line with an attention-driven mechanism, the third constitutes a confounding element. Unfortunately, I cannot control for this factor, since I do not have information on the semantic level.

Column (1) in Table 5 shows a descriptive analysis of the raw relationship between a link's position and the resulting average attention spillover: Articles are sorted into quintiles, which account for the order in which links appear. The first quintile are among the first 20% of the linked neighbors. Quintile 3 are articles in the middle and the last quintile contains articles which were linked at the end of the outgoing page. The coefficients show the average clicks on the articles in each quintile.[53]

The results suggest that articles appearing early in the treated (L_0) articles get a larger share of the additional attention than the articles in the middle and the lower part.[54] The first 20% of links receive much more attention than all other links. Specifically, they receive on average 88 more views than usual, whereas articles with links in the middle received about 60 additional views. The difference (28 views relative to qunitile 1) corresponds to a 32% smaller attention spillover for later links. Moreover, there is a weak but steady decline in attention from links in the middle to links that appear at the bottom, which would get only 53 additional views on average. This is less than 60% of the attention peak in the first quintile. These findings are robust to controlling for article features and aspects of the link structure (Column (2)) and to alternative measures of link position.[55]

[53] I consider the first appearance of a link, whenever it appears more often than once.

[54] I applied an outlier-correction(removing .5% extreme observations of each dependent variable).

[55] I used various alternative specifications. Results are also confirmed when employing specifications with alternative measures of link position, such as the precise position of a link (rather than quintile-bins;

Table 5: Properties of shocked page and relationship to spillover. Dependent variable: Peak of clicks on day 0.

	link order epicenter		clickshock vs. number of links		
	(1) raw correlation	(2) with controls	(3) L0 peak	(4) total L0 links	(5) Ratio L0 clicks/link
Dummy: Link among second 20% in L0	-31.547***	-25.144***	-24.698***	-24.685***	-24.272***
	(6.225)	(5.916)	(5.878)	(5.809)	(5.853)
Dummy: Link among third 20% in L0	-29.098***	-17.887***	-17.724***	-19.823***	-17.622***
	(6.314)	(5.996)	(5.968)	(5.906)	(5.941)
Dummy: Link among fourth 20% in L0	-33.864***	-23.026***	-22.737***	-23.411***	-22.470***
	(6.049)	(5.673)	(5.648)	(5.593)	(5.626)
Dummy: Link among fifth 20% in L0	-35.496***	-25.988***	-25.803***	-25.790***	-25.202***
	(6.083)	(5.886)	(5.865)	(5.782)	(5.833)
frequency of link's appearance in L0		15.572***	15.476***	15.700***	15.282***
		(3.695)	(3.659)	(3.708)	(3.628)
# links from other neighbors (L1)		0.508	0.501	1.495**	0.581
		(0.674)	(0.671)	(0.683)	(0.673)
Number of Revisions		0.186***	0.178***	0.185***	0.171***
		(0.055)	(0.055)	(0.054)	(0.055)
Number of authors		0.390**	0.419**	0.366**	0.436**
		(0.184)	(0.183)	(0.182)	(0.183)
Length of page (in bytes)		0.000	0.000	0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)
Number of images		-0.187	-0.172	-0.227	-0.170
		(0.309)	(0.303)	(0.303)	(0.303)
Clicks on L0 article			0.001***		
			(0.000)		
Number of links in L0				-0.111***	
				(0.007)	
Clicks per link on L0					0.233***
					(0.052)
Constant	88.122***	28.501***	24.598***	61.566***	22.698***
	(5.002)	(6.768)	(6.764)	(7.175)	(6.727)
Observations	6300	6300	6300	6300	6300
Number of Pages					
Adj. R^2	0.008	0.093	0.095	0.118	0.100

NOTES: The table shows how a neighbor's peak in traffic is related to the link's position and to different other properties of the shocked page (the peak in viewership on the treated article contrasted with the number of outgoing links on the page.) Columns (1)-(2) show the coefficients for link order. Columns (3)-(5) include the additional characteristics. Specification (3) gives the relationship to the clicks on the shocked page; (4) shows the association with the outgoing links. In Columns (5) I add the ratio of clicks per outgoing link on the shocked (L_0) page. The unit of observations is the outcome of a page i on day 0, the day of treatment of the treated (L_0) pages. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; no. of obs. = 6,300; no. of clusters = 82; no. of articles = 6,300.

Another source of heterogeneity is the fact that the advertised featured pages differ in how much attention they attracted. I analyze how the size of the initial spike (Column 3) and the number of competing links on the advertised article (Column 4) influence the additional attention that neighbors get. The analysis highlights that both factors matter for the average spillover with the expected signs. In column (5) I check if the number of clicks *per* link on the outgoing article is the best predictor for the average additional views on the neighbors. While this variable is relevant, it does not add predictive power.

I conclude this subsection by analyzing the association of link position to activity and to edit length. It is shown in Table 11 and suggests that superficial edits are influenced by the order in which links appear, but long edits are not. Columns (1) and (4) show the same regression as in the first column of Table 5. However, now the dependent variable is an indicator of activity (Columns 1-3) and the absolute change in article length (Columns 4-6). Edit activity is much more likely on an article linked at the top of the advertised neighbor (Column 1). Yet, Column (4) shows that the edit length does not depend on the link position (except for links in the fourth quintile). These findings are robust to controlling for page and network characteristics (Columns 2, 3, 5 and 6). These results support the interpretation of attention being a driver of activity, but not of edit depth.

6.3 Long-Run Content Improvements

I now turn to the long-run effects of the treatments that generate the exogenous variation to the system. Previous sections have highlighted that advertisement of a page results in increased attention and contribution activity to the neighboring pages. Hence, I quantify the "long-run" changes to content due to the advertisement. I aggregate the changes to the content that were present a week after the advertisement. Some changes will occur with a lag, if they fix flaws or content holes, that are not easily fixed right away. Looking at the difference in content a week after treatment captures also these (potentially more substantial) edits. Moreover, it helps to disentangle activity from improvements to the text. Changes which do not improve the article are usually reverted. This generates only activity but no improvement. However, an improvement will stay in the text and can be measured by taking the weekly difference.[56]

The findings are presented in Table 12. The first five columns show the changes that were still present on the treated page after a week. Columns (6)-(10) show the weekly changes aggregated over all neighbors. In both groups, the dependent variables are: (1) new authors, (2) change in length, (3) change in references, (4) change in images and (5) change in external links. All changes in content are measured in absolute terms. I

available upon request.).

[56]This analysis captures only changes in the absolute length or number of images/references etc. In further research I hope to also quantify quality improvements, that leave the absolute length untouched.

show difference-in-differences regressions, where the first difference is between the third observed week under real treatment vs. the third week under placebo treatment.[57] The second difference is between treated (clusters of) articles and those in the control group.

I find a very robust effect on the content of directly treated articles. On average they changed 285 bytes more than usually. Moreover, many references, external links and even images are modified and not reverted. This is harder to show for the aggregated activity over neighbors. Only the coefficients of the number of edits (not reported) and authors show a statistically significant increase in growth after the advertisement. However, for changes in content, references or pictures the coefficients fail to reject the null-hypothesis of no additional content growth. This is most striking for modified content ("page length"). While the point estimate is three to four times greater than the for direct treatment, the variance in aggregate contributions is too large to result in significant coefficients.[58] For the other content measures I find smaller and insignificant point estimates. This may mean that the contributions are small, but it may also mean that rather the quality than the quantity is improved. It might also just be due to aggregating contributions over the entire article clusters (and a whole week). The resulting variance may be more important, than the contributions due to treatment.

7 Discussion and Further Research

My analysis of the link network's causal effects reveals a robust pattern: links between articles channel attention which, in turn, drives contributions. Neighbors of shocked articles received 32 more visits on average - an increase of almost 70 percent. Moreover, I find that on average 1,000 views are needed before additional content (an edit) is generated. While the underlying spillover effect in clicks is substantial, the conversion to content is much smaller - even huge shocks did not generate many revisions on neighboring articles. This suggests that placing links strategically will only generate large effects if the pages that link out are very frequented. However, for the normal traffic on a typical Wikipedia page we would expect very small effects.

On the one hand my findings provide robust evidence for the ability of citation networks for channeling human *attention*. If this finding is generally valid, scientific citations could be used for strategic purposes, e.g. to raise awareness to related articles in other fields. On the other hand, my results suggest that using the link network is an expensive and inefficient strategy for channeling *contribution* flows in peer production, at least in

[57] Since I extract the two weeks before and after treatment, the third week in my dataset, is the first week after treatment.

[58] The point estimate suggests that over the cluster, 1000 bytes more of content than usual are change. I obtain very similar point estimates over several other specifications that use different reference groups or different ways to take difference-in-differences.

a mature wiki. Many users only look up information. Previous research suggested that smaller platforms (or offline communities) need other drivers of public goods contributions (e.g. social image and altruism (e.g. Carpenter and Myers (2010)). I can isolate the effect of attention from other determinants of public good contribution such as reputation, social image and altruism. Still, my findings confirm that attention-driven contributions alone will unlikely ensure sufficient provision of public goods in these settings.

These results are directly relevant for setting up a firm wiki or realizing the Wikimedia Foundation's vision.[59] My results further highlight that two out of three users click on a link. Possibly these estimates apply only to a mature wiki like the German Wikipedia. On the one hand, more studies could easily investigate this question using smaller or younger wikis. On the other hand, for interpreting my results in the context of knowledge-based peer production or a citation network using data from a mature wiki is advantageous. The findings suggest that links are being followed. The attention to a certain field or project will be higher if it receives links from other articles in other areas. While it will take more to also drive up production, my results suggest that policy makers should consider encouraging cross citations between fields.

My structural estimates suggest that an article will receive 22 percent of the average views on neighboring articles. Placing links to oft frequented nodes and thus increasing the average daily views on their neighbors by ten, one could obtain 2.2 additional daily visits to an article. I show that upper and lower bounds for the structural parameters can be computed even if the underlying network structure is unknown and illustrate how to compute them for my application. More generally, these bounds are easily computed for settings where only one node is treated in each subpopulation. I conjecture that they can easily be generalized to treatments that affect more than one node. This would allow nesting two-layered randomized control settings that aim at identifying social effects through exogenous variation over subpopulations.[60] It is thus no longer necessary to neglect the network structure in such experiments, merely because the information on links is not available. Moreover, the modeling approach is similar to spatial Durbin models (cf. LeSage (2008)). Hence, "mini population treatments" might be useful in spatial econometrics if the exogeneity of the spatial structure is in doubt.

Like with any other strategy, there are several limitations to the presented approach. Most importantly, I assume that neighbors of the treated nodes do not adjust their outcome as a reaction to the mere fact that their neighbor was treated. A counter example would be a teacher who selectively punishes or favors a single student: other pupils might react to the special treatment, e.g. become less motivated to study for the subject. Then their performance change reflects the sum of the spillover and their behavioral adjustment. In Appendix F I outline such a case and illustrate formally why the spillover parameter can

[59] A world where all "can freely share in the sum of all knowledge" (Wikimedia-Foundation (2013))

[60] Moffitt (2001), Angelucci and De Giorgi (2009), Kuhn et al. (2011), Crépon et al. (2013) etc.

no longer be identified. Another limitation is the assumption that the network formation *process* is not affected by the treatment. This is an imperfection in the match between my application and the methods for measuring attention spillovers. While the shocks on individual nodes and the high frequency are an asset, small content changes and small changes to the network occur in Wikipedia. The fact that they are limited leads me to consider the assumptions acceptable for Wikipedia's "today's featured article" but less so in other settings. Generally, if the content process is affected by treatment all estimates of indirect treatment effects will reflect a sum of the treatment on the existing network and new spillovers due to the changes in the link network. This might lead to upward biases (cf. Comola and Prina (2013)). Further research could exploit such exogenous shocks precisely as a trigger for link formation.

A promising area for further analysis would investigate whether the new contributions, especially the ones by new authors, add substantive knowledge or rather focus on improving small details. Future research should also exploit the heterogeneity in intensity of direct treatment effects more thoroughly. In particular, it would be interesting to analyze how attention, here measured as average effect, is distributed across neighbors. Is it evenly distributed or do users herd to only a few of the linked pages? Another promising area would use the methodology based on exogenous local treatments alongside that based on the network structure and the exploitation of open triads (Bramoullé et al. (2009), De Giorgi et al. (2010)). The approaches are complementary; research along these lines will result in valuable insights. Finally, it was not yet possible to surmount the computational hurdle of exploiting the detailed network information when obtaining the structural estimates. Future research should include this information and investigate which population parameter should be optimally included for relating reduced form and structural parameters. Also the analysis in section 6 highlights room for further research. Specifically, columns (3) - (5) of Table 5 highlight an important and well-understood trade-off that the designer of a network faces. While it is useful to be linked to a well frequented node, it is less helpful to be only one among many connections. Further, possibly experimental, research could analyze these issues in more detail. Finally, it is a worthy challenge to incorporate all moderating factors which I can only superficially explore here, into a fully fledged theoretical framework of spillovers.

8 Conclusions

I investigate how the link network between articles influences attention and subsequent content generation on German Wikipedia. I exploit exogenous short term shocks to attention and show how treatments diffuse across networks. Moreover, I can show how much content is generated due to additional attention. To the best of my knowledge these results are the first to quantify how a citation network causally influences users' contributions in an important online network of content pages. My analysis of the link network's causal effects reveals a robust pattern: citations channel attention which, in turn, drives contributions. Neighbors of shocked articles received almost 70 percent more visits and 1,000 views translate into 1 edit.

To uncover the underlying structural spillover, I augment the workhorse model for estimating peer effects (or spillovers) in networks (Bramoullé et al. (2009)). I incorporate exogenous treatments of individual nodes, which serve as complementary source of identification of the structural spillover effect. This allows relaxing the assumption of an exogenous network structure. I find that an article receives 22 percent of the average views on neighboring articles. I show that upper and lower bounds for the structural parameters can be computed even if the underlying network structure is unknown. These bounds, which are easily computed, allow accounting for the network structure in similar treatment control studies where the links are unobserved.

Several implications can be derived for knowledge based peer production, for setting up a firm wiki or for Wikipedia. Two thirds of all additional views translate into additional views on a neighbor. The significance of this result deserves emphasis: On average, *two out of three* visitors of "Today's featured article" click on a link to acquire further information. This large effects of links encourages the strategic use of citations to channel attention in other contexts of knowledge-based peer production. Moreover, the spillover does not depend on the targets' characteristics - all contents have a fair chance of getting some attention. However, what happens once the attention is there *does* depend on the items' characteristics - much less content is generated on *shorter* pages. Taken together, my findings suggest that (i) superficial edits are influenced by the citation network, but that deep edits are not. Moreover, (ii) only when many users contribute, attention-driven contributions alone may ensure sufficient provision of public goods.

References

Angelucci, Manuela and Giacomo De Giorgi, "Indirect Effects of an Aid program: How do Cash Transfers Affect Ineligibles' Consumption?," *The American Economic Review*, 2009, *99* (1), 486–508.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2008.

Aral, Sinan and Dylan Walker, "Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks," *Management Science*, 2011, *57* (9), 1623–39.

Ashenfelter, Orley and Michael Greenstone, "Using Mandated Speed Limits to Measure the Value of a Statistical Life," *Journal of Political Economy*, 2004, *112* (S1), S226–S267.

Ballester, Coralio, Antoni Calvo-Armengol, and Yves Zenou, "Who's Who in Networks. Wanted: The Key Player," *Econometrica*, 09 2006, *74* (5), 1403–17.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson, "The Diffusion of Microfinance," Technical Report, National Bureau of Economic Research 2012.

Barber, Brad M and Terrance Odean, "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors," *Review of Financial Studies*, 2008, *21* (2), 785–818.

Benkler, Yochai, "Coase's Penguin, or, Linux and" The Nature of the Firm"," *Yale Law Journal*, 2002, pp. 369–446.

_ , *The wealth of networks: How social production transforms markets and freedom*, Yale Univ Pr, 2006.

_ **and Helen Nissenbaum**, "Commons-based Peer Production and Virtue*," *Journal of Political Philosophy*, 2006, *14* (4), 394–419.

Berger, Jonah, Alan T Sorensen, and Scott J Rasmussen, "Positive effects of negative publicity: When negative reviews increase sales," *Marketing Science*, 2010, *29* (5), 815–827.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 2009, *150* (1), 41–55.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, "Peer effects and social networks in education," *The Review of Economic Studies*, 2009, *76* (4), 1239–1267.

Carmi, Eyal, Gal Oestreicher-Singer, and Arun Sundararajan, "Is Oprah Contagious? Identifying Demand Spillovers in Online Networks," *NET Institute Working Paper.*, 2012, *No. 10-18.*, Available at SSRN: http://ssrn.com/abstract=1694308 or http://dx.doi.org/10.2139/ssrn.1694308 (August 3, 2012).

Carpenter, Jeffrey and Caitlin Knowles Myers, "Why volunteer? Evidence on the role of altruism, image, and incentives," *Journal of Public Economics*, 2010, *94* (11), 911–920.

Claussen, Joerg, Oliver Falck, and Thorsten Grohsjean, "The Strength of Direct Ties: Evidence from the Electronic Game Industry," *International Journal of Industrial Organization*, 2012, *30* (2), 223–30.

Coase, Ronald H, "The nature of the firm," *economica*, 1937, *4* (16), 386–405.

Comola, Margherita and Silvia Prina, "The Impact of Exogenous Interventions on Networks: A Dynamic Peer Effect Model," 2013.

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora, "Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment," *The Quarterly Journal of Economics*, 2013, *128* (2), 531–80.

Dahl, Gordon B, Katrine V Løken, and Magne Mogstad, "Peer Effects in Program Participation," Technical Report, National Bureau of Economic Research 2012.

Dellarocas, Chrysanthos, Zsolt Katona, and William Rand, "Media, Aggregators, and the Link Economy: Strategic Hyperlink Formation in Content Networks," *Management Science*, 2013, *59* (10), 2360–2379.

Ferracci, Marc, Grégory Jolivet, and Gerard J. van den Berg, "Evidence of Treatment Spillovers within Markets," online 2012. last retrieved on Nov. 11, 2013 at http:www.efm.bris.ac.ukecgrjeval_resubmit.pdf.

Fershtman, Chaim and Neil Gandal, "Direct and Indirect Knowledge Spillovers: The "Social Network" of Open-source Projects," *RAND Journal of Economics*, 2011, *42* (1), 70–91.

Giorgi, Giacomo De, Michele Pellizzari, and Silvia Redaelli, "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, 2010, *2* (2), 241–75.

Giorgidze, George, Torsten Grust, Tom Schreiber, and Jeroen Weijers, "Haskell Boards the Ferry: Database-supported Program Execution for Haskell," in "Revised selected papers of the 22nd international symposium on Implementation and Application of Functional Languages, Alphen aan den Rijn, Netherlands," Vol. 6647 of *Lecture Notes in Computer Science* Springer 2010. Peter Landin Prize for the best paper at IFL 2010.

Graham, Brian S., "Methods of Identification in Social Networks," *Annu. Rev. Econ.*, 2015., *7*, Submitted. Doi: 10.1146/annurev-economics-080614-115611.

Hinz, Oliver, Bernd Skiera, Christian Barrot, and Jan U Becker, "Seeding strategies for viral marketing: an empirical comparison," *Journal of Marketing*, 2011, *75* (6), 55–71.

Ho, Jason YC and Melanie Dempsey, "Viral marketing: Motivations to forward online content," *Journal of Business Research*, 2010, *63* (9), 1000–1006.

Hoffman, Andrew J and William Ocasio, "Not all events are attended equally: Toward a middle-range theory of industry attention to external events," *Organization Science*, 2001, *12* (4), 414–434.

Hu, Nan, Yi Dong, Ling Liu, and Lee J Yao, "Not all that glitters is gold: the effect of attention and blogs on investorsŠ investing behaviors," *Journal of Accounting, Auditing & Finance*, 2013, *28* (1), 4–19.

Huberman, Bernardo A, Daniel M Romero, and Fang Wu, "Crowdsourcing, attention and productivity," *Journal of Information Science*, 2009, *35* (6), 758–765.

Imberman, Scott, Adriana D Kugler, and Bruce Sacerdote, "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees," Technical Report, National Bureau of Economic Research 2009.

Jackson, Matthew O. and Yves Zenou, *Economic Analyses of Social Networks, The International Library of Critical Writings in Economics*, London: Edward Elgar Publishing, 2013.

Jacobson, Louis S, Robert John LaLonde, and Daniel G Sullivan, "Earnings Losses of Displaced Workers," *American Economic Review*, 1993, *83* (4), 685–709.

Keegan, Brian, Darren Gergle, and Noshir Contractor, "Hot Off the Wiki Structures and Dynamics of Wikipedia's Coverage of Breaking News Events," *American Behavioral Scientist*, 2013, *57* (5), 595–622.

Kuhn, Peter, Peter Kooreman, Adriaan Soetevent, and Arie Kapteyn, "The Effects of Lottery Prizes on Winners and their Neighbors: Evidence from the Dutch Postcode Lottery," *The American Economic Review*, 2011, *101* (5), 2226–47.

Kummer, Michael E, Marianne Saam, Iassen Halatchliyski, and George Giorgidze, "Centrality and Content Creation in Networks – The Case of German Wikipedia," *ZEW-Centre for European Economic Research Discussion Paper*, 2012, (12-053).

Lerner, Josh and Jean Tirole, "Some simple economics of open source," *Journal of Industrial Economics*, 2002, pp. 197–234.

LeSage, James P, "An introduction to spatial econometrics," *Revue d'économie industrielle*, 2008, (3), 19–44.

Manski, Charles F, "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 1993, *60* (3), 531–42.

Mayzlin, Dina and Hema Yoganarasimhan, "Link to success: How blogs build an audience by promoting rivals," *Management Science*, 2012, *58* (9), 1651–1668.

Miguel, Edward and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, 2003, *72* (1), 159–217.

Moffitt, Robert A, "Policy Interventions, Low-level Equilibria, and Social Interactions," *Social Dynamics*, 2001, pp. 45–82.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd, "The PageRank citation ranking: Bringing order to the web.," 1999.

Sacerdote, Bruce, "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 2001, *116* (2), 681–704.

Shang, Jen and Rachel Croson, "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods," *The Economic Journal*, 2009, *119* (540), 1422–1439.

Wikimedia-Foundation, "Frequently Asked Questions," online 2013. last retrieved on Sept. 9, 2013 at http://wikimediafoundation.org/wiki/FAQ/en.

Wu, Fang and Bernardo A Huberman, "Novelty and collective attention," *Proceedings of the National Academy of Sciences*, 2007, *104* (45), 17599–17601.

Zhang, Xiaoquan and Feng Zhu, "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia," *The American Economic Review*, 2011, *101*, 1601–15.

A Data-Appendix

Table 6: Summary statistics: direct neighbors of shocked articles in the 'featured articles' condition

	mean	sd	min	p10	p50	p90	max
Length of page (in bytes)	6,686	6,811	16	65	4,677	15,033	111,000
Number of authors	31	34	0	2	20	75	324
Clicks	35	147	0	0	0	82	25,574
Number of Revisions	92	132	0	4	45	231	2,106
Links from Wikipedia	117	315	0	6	34	266	10,131
Dummy: literature section	.3	.46	0	0	0	1	1
Number of images	1.9	6	0	0	1	4	319
Number language links	14	19	0	0	7	41	180
References (footnotes)	1.3	4.1	0	0	0	3	182
Links to further info	2.3	4.5	0	0	1	6	200
time variable (normalized)	0	8.4	-14	-12	0	12	14
Delta: Number of Revisions	.042	.38	0	0	0	0	42
Delta: Length of page	2.1	201	-91,137	0	0	0	99,782
Delta: Number of authors	.014	.13	0	0	0	0	13
Delta: Links from Wikipedia	.061	1	-90	0	0	0	473
Delta: Number of images	.00054	.19	-76	0	0	0	132
Delta: References	.0014	.18	-104	0	0	0	57
Delta: Links further info	.0008	.12	-25	0	0	0	50

NOTES: The table shows the distribution of the main variables. The unit of observations is the outcome of a page i on day t. The time variable is normalized and runs from -14 to 14.; no. of obs. = 908,628; no. of start pages = 174; no. of articles = 15,732.

Table 7: Advertised “featured articles”, associated neighbors and corresponding observations.

name of event	pages	observations		
	no.	control	treated	Total
Afrikaans	128.0	5,481.0	1,943.0	7,424.0
Alte_Synagoge_(Heilbronn) [old synagogue of Heilbronn]	52.0	1,885.0	1,131.0	3,016.0
Banjo-Kazooie	125.0	5,191.0	2,030.0	7,221.0
Benno_Elkan	139.0	5,133.0	2,900.0	8,033.0
Bombardier_Canadair_Regional_Jet	92.0	4,205.0	1,073.0	5,278.0
CCD-Sensor	586.0	31,001.0	2,871.0	33,872.0
Charles_Sanders_Peirce	258.0	11,716.0	3,219.0	14,935.0
Das_Kloster_der_Minne [the monastery of courtly love]	51.0	1,827.0	1,102.0	2,929.0
Deutsche_Bank ["German bank"]	343.0	10,005.0	9,860.0	19,865.0
Eishockey [ice hockey]	162.0	4,698.0	4,698.0	9,396.0
Ekel [disgust]	270.0	10,295.0	5,336.0	15,631.0
Fahrbahnmarkierung [road surface marking]	44.0	1,276.0	1,276.0	2,552.0
Geschichte_Ostfrieslands [history of East Frisia]	235.0	7,453.0	6,177.0	13,630.0
Geschichte_der_deutschen_Sozialdemokratie [History of the social democratic party of Germany]	306.0	9,599.0	8,033.0	17,632.0
Glanzstoff_Austria	270.0	14,094.0	1,537.0	15,631.0
Glorious_Revolution	153.0	6,206.0	2,668.0	8,874.0
Granitschale_im_Lustgarten [granite cup in the pleasure garden (Berlin)]	83.0	3,857.0	928.0	4,785.0
Gustav_Hirschfeld	142.0	6,438.0	1,740.0	8,178.0
Hallenhaus [“Aisled house” (low German house)]	71.0	2,117.0	2,001.0	4,118.0
Helgoland	228.0	8,120.0	5,104.0	13,224.0
Jaroslawl [Yaroslavl]	321.0	12,789.0	5,829.0	18,618.0
Jupiter_und_Antiope_(Watteau) [Jupiter and Antiope (Watteau)]	36.0	1,160.0	928.0	2,088.0
Karolingische_Buchmalerei [Carolingian illumination]	162.0	4,843.0	4,553.0	9,396.0
Katholische_Liga_(1538) [Catholic league (1538)]	37.0	1,682.0	464.0	2,146.0
Martha_Goldberg	55.0	1,595.0	1,595.0	3,190.0
Naturstoffe [natural product]	320.0	9,338.0	9,222.0	18,560.0
Paul_Moder	61.0	1,798.0	1,682.0	3,480.0
St._Martin_(Memmingen) [Saint Martin church (Memmingen)]	59.0	1,653.0	1,711.0	3,364.0
Stabkirche_Borgund [Borgund stave church]	40.0	1,421.0	899.0	2,320.0
Taiwan	167.0	5,017.0	4,669.0	9,686.0
USS_Thresher_(SSN-593)	90.0	3,712.0	1,479.0	5,191.0
Visum [visa (document)]	56.0	1,624.0	1,624.0	3,248.0
Wenegnebti	55.0	1,798.0	1,363.0	3,161.0
Werder_Bremen	292.0	8,555.0	8,323.0	16,878.0
Total	5,489.0	207,582.0	109,968.0	317,550.0

NOTES: The table shows the “featured articles” from the in the dataset. For reasons of space, I show only advertisements on the 10^th of each month. Column 1 shows the number of associated articles that are one click away from one of the corresponding start pages (be it treated or control). Columns 2-4 show the number of observations. Observations associated with actually treated articles are shown separately from control observations. Pages can be accessed by pasting the title behind the last slash in: *http://de.wikipedia.org/wiki/*

A.1 Comparison of the Trends

Table 8: Summary statistics of variables' first differences. Comparison groups

	count	mean	sd	min	p10	p50	p90	max
Clicks	100212	32	121	0	0	0	74	9683
Delta: Number of Revisions	93054	.039	.34	0	0	0	0	20
Delta: Length of page	93054	1.9	186	-31473	0	0	0	31462
Delta: Number of authors	93054	.013	.12	0	0	0	0	6
Delta: Links from Wikipedia	93054	.057	.75	-58	0	0	0	124
Delta: Number of images	93054	.00074	.066	-3	0	0	0	12
Delta: References	93054	.00089	.07	-5	0	0	0	9
Delta: Links further info	93054	.00058	.073	-10	0	0	0	5

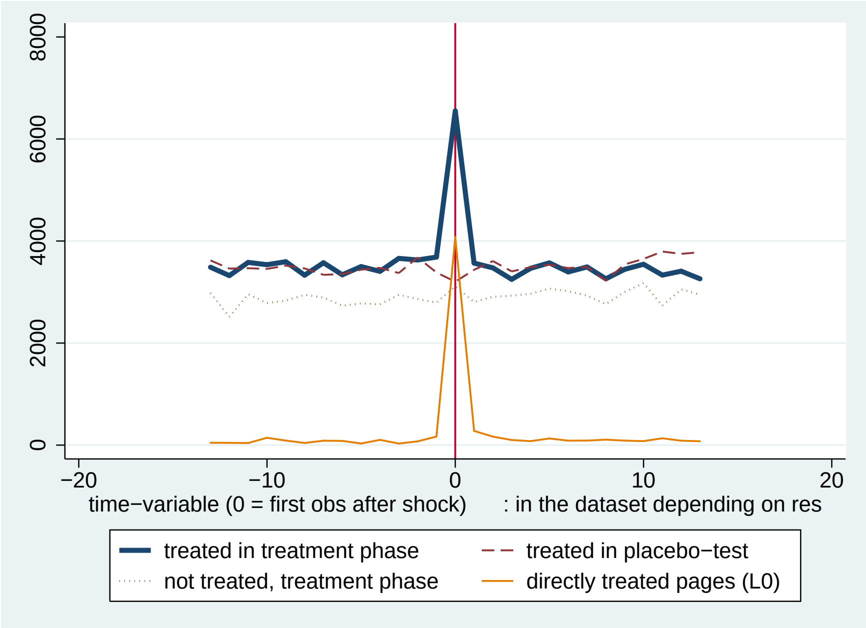
Table 9: Summary statistics of variables' first differences. Treated group before treatment.

	count	mean	sd	min	p10	p50	p90	max
Clicks	53088	34	135	0	0	0	83	10427
Delta: Number of Revisions	49296	.049	.48	0	0	0	0	42
Delta: Length of page	49296	2.6	125	-11666	0	0	0	11359
Delta: Number of authors	49296	.017	.15	0	0	0	0	9
Delta: Links from Wikipedia	49296	.048	.75	-50	0	0	0	86
Delta: Number of images	49296	.0044	.62	-8	0	0	0	132
Delta: References	49296	.0017	.13	-18	0	0	0	9
Delta: Links further info	49296	.00099	.13	-19	0	0	0	10
time variable (normalized)	53088	-7.5	4	-14	-13	-7.5	-2	-1

B Additional Regressions, Robustness and Figures

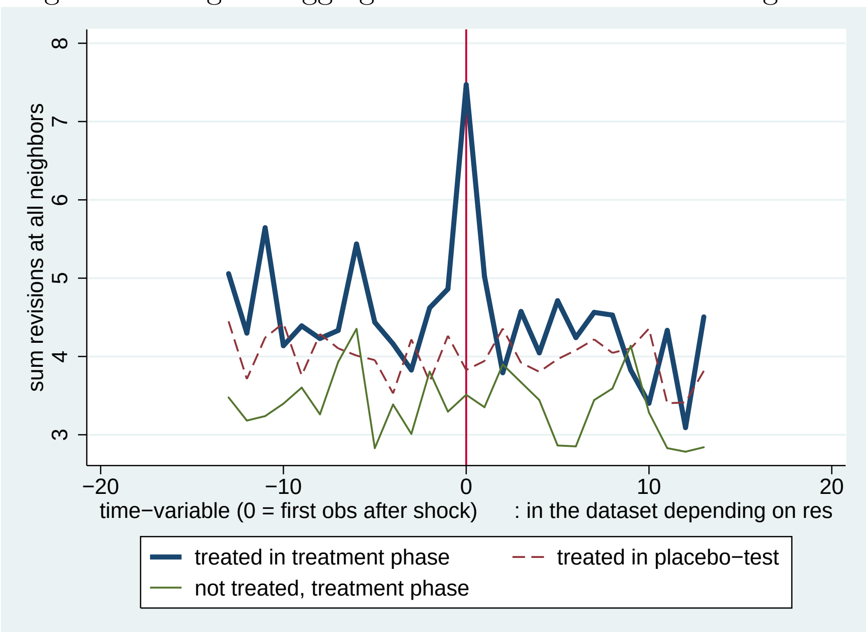
B.1 Additional Results - Aggregated Effects

Figure 4: Figure contrasting the mean of clicks on featured articles, with the aggregated clicks on all neighboring pages.



Notes: The figure shows the aggregated effect on the pages that are one click away. The average treated page received up to 4000 additional clicks, all neighbors together received approx. the same number of additional clicks

Figure 5: Figure showing the aggregated new revisions on all neighboring pages.



Notes: The figure shows the aggregated effect on the pages that are one click away. All neighbors of treated articles together received approx. three additional revisions.

B.2 Additional Results - Heterogeneities

B.2.1 Conversion to Edits

Table 10 shows how article characteristics influence the spillovers. In the first three columns of Table 10 I added variables that account for an article's length, how well it is generally linked and how closely they are linked to the shocked articles (by counting closed triads with other neighbors). Columns 4-6 show the results of an analysis that considers short articles ("stubs") seperately.

B.2.2 Link Position

The Table shows the download frequency for articles grouped by their order of appearance in the treated article. It contrasts different sets of control variables and two dependent variables (A Dummy which captures any activity and the abolute length of edits). The finding of columns 1 and 4 is very robust, as I document in the online appendix. There, more specifications for the robustness of the results for link position corroborate that attention to the first 20% of the links is significantly higher than attention to other articles.

B.2.3 Size of the Primary Effect and Spillovers.

I further analyzed (analysis not shown in the paper) how the amount of additional attention (i.e.: estimated number of clicks due to treatment) to the shocked pages (L_0) is related to the size of the attention spillover (=clicks) on the neighbors (L_1). I find a positive baseline effect, indicating that more users arrive at the average neighbor if the initial shock is large. Moreover, the positive effect is reduced for short articles and amplified for long and well linked articles. These results can be found in the online appendix.

Table 10: Relationship of clicks/added revisions and time dummies, including article heterogeneity.

	joint estimation			short articles only		
	(1) clicks	(2) revisions	(3) length	(4) clicks	(5) revisions	(6) length
t = -2	1.284	-0.003	3.028	2.105	0.001	0.493
	(2.833)	(0.008)	(2.066)	(2.317)	(0.003)	(0.334)
t = -1	2.720	0.005	-0.441	8.041*	-0.006	0.009
	(2.839)	(0.006)	(1.298)	(4.769)	(0.006)	(0.368)
t = 0	32.058***	0.042***	1.214	25.234***	0.012**	1.461*
	(6.757)	(0.010)	(1.578)	(6.781)	(0.005)	(0.744)
t = 1	-0.193	0.013*	2.486	4.372	-0.002	-0.146
	(1.852)	(0.008)	(2.845)	(4.138)	(0.001)	(0.498)
t = 2	-1.480	-0.007	0.276	3.971	0.000	0.054
	(2.301)	(0.007)	(1.187)	(4.518)	(0.005)	(0.365)
t = 3	-3.388	-0.002	-2.617	-3.383	-0.003	0.808
	(2.800)	(0.007)	(1.925)	(2.051)	(0.004)	(0.569)
t = 4	-1.628	-0.005	3.608	7.102	-0.003*	0.265
	(2.733)	(0.006)	(4.082)	(5.033)	(0.002)	(0.318)
Short article on t = 0	-6.450*	-0.029***	0.221			
	(3.899)	(0.007)	(1.221)			
Short article on t = 1	-0.547	-0.013**	-2.684			
	(1.441)	(0.006)	(2.116)			
Short article on t = 2	-0.410	0.004	0.372			
	(1.253)	(0.006)	(0.640)			
Many L1 links article on t = 0	33.339*	0.013	2.572			
	(17.370)	(0.017)	(4.256)			
Many L1 links article on t = 1	2.214	-0.002	7.807			
	(4.356)	(0.015)	(7.134)			
Many L1 links article on t = 2	-0.947	0.006	6.425**			
	(4.209)	(0.018)	(2.692)			
Well-linked article on t = 0	-19.404**	-0.031**	-2.079			
	(9.271)	(0.015)	(2.872)			
Well-linked article on t = 1	1.128	-0.013	-0.676			
	(2.971)	(0.012)	(4.670)			
Well-linked article on t = 2	0.425	0.004	-1.051			
	(3.417)	(0.010)	(1.857)			
Long article on t = 0	1.459	0.004	-0.139			
	(7.744)	(0.016)	(4.517)			
Long article on t = 1	4.404	0.000	-1.434			
	(2.986)	(0.016)	(4.807)			
Long article on t = 2	2.702	-0.034***	-7.764***			
	(3.611)	(0.012)	(2.684)			
Crossterms	Yes	Yes	Yes	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	36.208	0.045	2.971	13.285	0.004	0.163
Observations	346104	346104	346104	58080	58080	58080
Number of Pages	15732	15732	15732	2640	2640	2640
Adj. R^2	0.003	0.000	-0.000	0.002	0.000	0.000

Standard errors in parentheses
Fixed Effects Panel-Regressions with heteroscedasticity robust and clustered standard errors.
Only crossterms closer to treatment are shown, but all were included. Reference group t-14 to t-5.
* p<0.10, ** p<0.05, *** p<0.01

Table 11: Effect of the outgoing link's position in shocked page on activity and long edits.

	Dummy: Activity			Absolute Change in Page Length		
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy: Link among second 20% in L0	-0.019***	-0.019***	-0.017**	0.047	0.030	0.070
	(0.007)	(0.007)	(0.007)	(0.094)	(0.092)	(0.097)
Dummy: Link among third 20% in L0	-0.016**	-0.016**	-0.014**	0.057	0.033	0.063
	(0.007)	(0.007)	(0.007)	(0.091)	(0.092)	(0.095)
Dummy: Link among fourth 20% in L0	-0.025***	-0.025***	-0.023***	-0.192**	-0.218**	-0.192**
	(0.007)	(0.007)	(0.006)	(0.090)	(0.091)	(0.096)
Dummy: Link among fifth 20% in L0	-0.016**	-0.016**	-0.013*	0.079	0.053	0.093
	(0.007)	(0.007)	(0.007)	(0.092)	(0.094)	(0.101)
Number of revisions		0.000	0.000		-0.000	-0.000
		(0.000)	(0.000)		(0.001)	(0.001)
Number of authors		0.000	0.000		-0.001	-0.002
		(0.000)	(0.000)		(0.002)	(0.003)
Length of page (in bytes)		0.001***	0.001**		0.002**	0.001**
		(0.000)	(0.000)		(0.000)	(0.000)
Number of images		-0.001**	-0.002**		-0.006	-0.008
		(0.000)	(0.000)		(0.013)	(0.012)
Links from Wikipedia		-0.002***	-0.001***		-0.001	-0.001**
		(0.000)	(0.000)		(0.000)	(0.000)
Short article			-0.011**			0.194
			(0.005)			(0.206)
Long article			-0.004			-0.142
			(0.009)			(0.098)
Well-linked article			-0.008			0.184**
			(0.005)			(0.085)
Many L1 links article			0.004			0.117
			(0.010)			(0.139)
Frequency of link's appearance in L0			0.008*			0.031
			(0.004)			(0.043)
Number of links in L0			-0.001***			-0.001*
			(0.000)			(0.000)
# links from other neighbors (L1)			0.001			-0.004
			(0.001)			(0.014)
Constant	0.040***	0.027***	0.035***	0.588***	0.610***	0.606***
	(0.006)	(0.006)	(0.009)	(0.056)	(0.074)	(0.114)
Observations	6318	6318	6318	256	256	256
Adj. R^2	0.00	0.01	0.01	0.02	0.04	0.05
Mean of Dep. Var.	0.02	0.02	0.02	0.58	0.58	0.58

NOTES: The Table shows the relationship of a link's position in the outgoing text and activity on the link's target. The dependent variable in the first three columns is an variable indicating any editing activity on the page. The second three columns regresses the length of the text change (additions and deletions) on the links position, conditional that any changes were made. The first column in each group includes only dummies that indicate the Quintile of the link's position, The second column includes article characteristics, the third also accounts for network characteristics. Columns 4-6 do not include the 2 percent most substantial changes to account for outliers. The first quintile represents the neighbors that are mentioned among the first 20% of links in the outgoing page. The third quintile are article in the middle and the last quintile was linked at the end of the outgoing page. ; standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$; no. of obs. = 6,318; no. of clusters = 82; no. of articles = 6,318.

B.3 Long Term Effects on Content Generation

Table 12: Longterm-effects on Content Generation on the treated page vs. all neighbors.

	changes on featured page					changes on all neighbors				
	(1) authors	(2) length	(3) refs	(4) imagelinks	(5) extlinks	(6) authors	(7) length	(8) refs	(9) imagelinks	(10) extlinks
Treated observation	7.090***	215.056**	6.010**	1.312***	1.783***	2.352***	993.757	1.470	0.213	0.589
	(0.483)	(86.320)	(2.387)	(0.414)	(0.652)	(0.764)	(1062.796)	(1.226)	(0.367)	(0.948)
Constant	-0.155*	2.869	0.012	-0.000	0.024	-0.143	233.964	0.310	0.321	0.679
	(0.085)	(14.770)	(0.021)	(.)	(0.041)	(0.546)	(281.849)	(0.411)	(0.285)	(0.430)
Number of Page Clusters	177	177	177	177	177	170	170	170	170	170

NOTES: The table shows compares the content that was generated and persisted for a week. The first five columns show the changes that were mode on the treated page. Columns (6)-(10) show the after week changes aggregated over all neighbors. In both groups, the dependent variables are: (1) new authors, (2) change in length, (3) change in references, (4) change in images and (5) change in extrenal links. All changes in content are measured in absolute terms. Diff in Diff regressions: The first difference is between the third week under real vs. the third week under placebo, The second difference is between treated (clusters of) articles and untreated ones. Heteroscedasticity robust standard errors. The unit of observation is the outcome page (or cluster) i. Standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$; no. of obs. = 340; no. of clusters = 170; no. of articles = 340.

C Details about Data Preparation, the Treated and the Control Groups

This section gives detailed information about the preparation and storage of the dataset. The subsequent subsections explain how the database was put together and the procedure I used to extract the dataset that I use.

C.1 Preparation and Extraction

The dataset is based on a full-text dump of the German Wikipedia from the Wikimedia toolserver. To construct the history of the articles' hyperlink network for the entire encyclopedia, it was necessary to parse the data and identify the links. From the resulting tables, I constructed a time-varying graph of the article network, which provided the foundation for how I sample articles in my analysis. Furthermore, information about the articles, such as the number of authors who contributed up to a particular point in time or the existence of a section with literature references was added. Hence, the data I use are based on 153 weeks of the the entire German Wikipedia's revision history between December 2007 and December 2010. Since the data are in the order of magnitude of terabytes, it was not be possible to conduct the data analysis using only in-memory processing. We therefore stored the data in a relational database (disk-based) and queried the data using Database Supported Haskell (DSH) (Giorgidze et al. (2010)).[61]

C.2 Choice of Treated Articles and Neighborhood

"Featured articles" were found by consulting the German Wikipedia's archive of pages that were selected to be advertised on Wikipedia's main page ("Seite des Tages") between December 2007 and December 2010. To reduce the computational burden and to avoid the risk of temporal overlaps of different treatments, I focus on pages that were selected on the 10^{th} of a month. I identified all the pages that received a direct link (L_1) or an indirect link (L_2) from such a featured article more than a week before treatment. I evalutated links with this time gap before the shock actually occurred to make sure that the results are not driven by endogeneous link formation.[62] Having fixed the set of pages to observe, I extracted daily information on the contemporary state of the articles (page visits, number of revisions, number of distinct authors that contributed, page length, number of external links etc.). I determine these variables on a daily basis, 14 days before the event occurred (on a neighboring page) and 14 days after the shock (giving a total of 29 observations per page).

I am most interested in attention spillovers and content provision, which are not directly related to the events but rather a consequence of the spike in interest and the resulting improvements to the linked pages. Hence, I do not focus on the treated pages directly, but on the set L_1 that are "one click away", in my analysis of the "featured articles".[63]

C.3 Choice of Control Group Articles and Neighborhood

The approach I take in this paper hinges on the availability of a valid control group. To obtain such observations I pursue two distinct strategies. The first approach uses pages which are similar but unlikely to be affected by the treatment. For a first comparison I selected other featured articles and neighbors thereof that were advertised featured either later or earlier in time. Given such a similar page, I identified their direct and indirect neighbors when the event occurred on the treated page. This gives me a set $C1_{control}$ which is similar in both size and characteristics to the sampled pages (before the shock). Yet, the choice of the start pages in the comparison group is somewhat arbitrary.[64] I address this issue by simulating a treatment

[61] This is a novel high-level language allowing the writing and efficient execution of queries on nested and ordered collections of data.

[62] I thus only include pages that had a link before it was known that the start page will be hit. I furthermore exclude pages that receive their indirect (L_2) link via a page that has more than 100 links, since such pages are very likely either pure "link pages" very general pages (such as pages about a year), that bare only a very weak relationship to the shocked site.

[63] Effects on the pages that are 2 clicks away were to small too be measured.

[64] Ideally the selection of comparison pages should be based on matching procedures, which is unfortunately not possible without computing the characteristics of all the 1,000,000 nodes. My approach is

on the treated pages 42 days before the event occurred. I refer to the articles in this "placebo-treatment" as $C1_{placebo}$, because for them $t = 0$ when no actual treatment occurred. By design, this comparison group consists of the same set of articles (treated and their neighbors). This comes at the cost of observing the articles at a different point in time. A third control group of "unrelated" observations results from applying a placebo to the control group.[65]

Table 7 (in the data appendix) shows a representative set of featured articles, which were chosen by my procedure and included in the data. In general, they cover various topics such as innovations (e.g. the CCD-sensor), places (Helgoland), soccer clubs (Werder Bremen) or art history topics (Carolingian book illustrations). The first column of the table shows the number of articles that belong to each featured article. The last three columns show the number of observations that received a link from an article before it was advertised featured, separated by whether or not they belong to a time-series with actually treated observations.[66] The numbers range from 2,088 to 33,872.

C.4 Extraction for Natural and Technical Disasters

In my most important robustness check (cf. Appendix D.2) I analyze a different set of shocks to attention that stem from accidents and natural disasters with sudden onset (e.g. earthquakes and plane crashes). To identify major events, I consulted the corresponding page on Wikipedia and selected the 26 largest events with spontaneous onset. For each of these events we identified the page that corresponds to the event, which are considered to be in the set "L_0" (sometimes also called "start pages"). Note that this page is typically created *after the event* occurred[67], which obliges me to identify the pages, that user will most likely turn to until the disaster's page is in place. To achieve this, I used the link data to identify the set of pages that later shared a reciprocal link with the start page. Such a reciprocal link indicates that they were closely related to the event. After the event page came in to existence they were only one click away (set "L_1"). Next, we identified those pages that received a link from an L_1 page (unidirectional) (2 clicks away set "L_2") For disasters the shock is very large and the event page usually does not exist at the time of the shock, so the L_1 pages might have been treated themselves.[68] Hence, I focus on the indirectly linked set of pages (L_2) in the analysis below.

For disasters I proceeded along similar lines. I focused on the network around older catastrophes that occurred at a different point in time and were not from exactly the same domain, to avoid overlaps in the link network ($C_{control}$). Alternatively, I observe the same set of pages seven weeks before the disaster ($C_{placebo}$). Table 17 shows which events were included in the data and shows the associated number of observations for each of them. The dataset includes both natural disasters as well as technical or economic catastrophes.

For sudden onset events the similarity in the trends of the two groups is even more striking: treatment and comparison groups have literally equal trends, underscoring the usefulness of this supplementary dataset.

C.4.1 Natural Disasters

however quite robust independently of how I specify the control group. I also compared to the neighbors around articles of similar size and relative importance, about similar topics, but in a remote geographic space or technical domain. Such a change in the specification of the control group does not affect my results. (available upon request).

[65]This set of observations actually emerged as an artifact from the data extraction. Nevertheless it provides yet another group that can be compared to the treated group.

[66]Note, that each page shows up 29 times in the raw data and was sampled twice (placebo and real treatment), so that the number of corresponding pages (treatment or control) can be inferred by dividing the number of observations by 58.

[67]Usually it takes up to two days until the event receives its own page.

[68]Some of the consequences of major events, such as earthquakes, might change the state of the world and thus trigger a change in content, which is merely *due to the event* (e.g. destruction of an important monument). Consequently, I do not emphasize the change in activity on the pages that are only one click away for disasters. I also exclude pages if they were later directly linked to the event page.

Table 13: Summary statistics of variables' first differences. Comparison groups

	count	mean	sd	min	p10	p50	p90	max
Clicks	356961	31	148	0	0	0	67	10733
Delta: Number of Revisions	344652	.033	.34	0	0	0	0	44
Delta: Length of page	344652	1.7	98	-20197	0	0	0	27500
Delta: Number of authors	344652	.012	.12	0	0	0	0	8
Delta: Links from Wikipedia	344652	.049	2.7	-1148	0	0	0	216
Delta: Number of images	344652	.00052	.068	-7	0	0	0	20
Delta: References	344652	.0012	.13	-32	0	0	0	29
Delta: Links further info	344652	.001	.12	-15	0	0	0	31

Table 14: Summary statistics of variables' first differences. Treated group before treatment.

	count	mean	sd	min	p10	p50	p90	max
Clicks	33320	32	210	0	0	0	71	15304
Delta: Number of Revisions	30940	.035	.32	0	0	0	0	14
Delta: Length of page	30940	1.9	190	-22416	0	0	0	22416
Delta: Number of authors	30940	.012	.12	0	0	0	0	3
Delta: Links from Wikipedia	30940	.051	.76	-37	0	0	0	59
Delta: Number of images	30940	-.00016	.16	-27	0	0	0	6
Delta: References	30940	.0016	.084	-4	0	0	0	8
Delta: Links further info	30940	.0014	.085	-6	0	0	0	4
time variable (normalized)	33320	-7.5	4	-14	-13	-7.5	-2	-1

Table 15: Summary statistics of variables' first differences. Treated group after treatment.

	count	mean	sd	min	p10	p50	p90	max
Clicks	33320	48	273	0	0	0	99	14160
Delta: Number of Revisions	33320	.055	.46	0	0	0	0	24
Delta: Length of page	33320	2.4	70	-4811	0	0	0	3042
Delta: Number of authors	33320	.021	.18	0	0	0	0	11
Delta: Links from Wikipedia	33320	.051	.64	-38	0	0	0	30
Delta: Number of images	33320	.00051	.059	-4	0	0	0	7
Delta: References	33320	.0025	.13	-5	0	0	0	13
Delta: Links further info	33320	.0018	.11	-5	0	0	0	13
time variable (normalized)	33320	7.5	4	1	2	7.5	13	14

D Robustness Checks

D.1 Large Events

In this section I describe a robustness check that repeats the estimation procedure for obtaining the main results on a different dataset that is based on 23 catastrophic events with sudden and unpredictable onset. Unlike "today's featured articles" these shock to attention and activity are beyond the control of the platform owners and are, by construction, harder to predict. Hence, using these data requires a different set of assumptions and my purpose is to show that the results are very similar, despite the changed setting. If the estimated spillovers are of the same magnitude, this would be reassuring evidence, that the selection mechanism by which featured articles are chosen to be advertised is not driving my results.

Here I describe the dataset and the results. More details about the data preparation and selection of the events are provided in Appendix C, together with more details on data extraction.

D.1.1 Large Events - Preparation and Compared Groups

To identify major events, I consulted the corresponding page on Wikipedia. The most important feature of major events is that they are arguably exogenous to Wikipedia and unpredictable to Wikipedians. However, "unpredictability" is threatened for events that take several days to build up (e.g. floods, huricanes or ash clouds produced by volcanos) and are hence predictable in the sense that experts might foresee the disastrous event before it is in the media. To avoid this problem, I focus on 26 large events with spontaneous onset, e.g. earthquakes and accidents. I focus on the content provision that results from attention spillovers and which is a consequence of the spike in interest and the resulting improvements to the linked pages. Hence, I will not focus on the treated pages where content generation might be related to the events directly. Instead I obtained data on the direct and indirect network neighbors. Table 17 shows the information for the data on large events which includes both natural disasters as well as technical or economic catastrophes.

D.1.2 Large Events - Descriptive Statistics

Summary statistics for the data on large events are shown in Table 16. The data contains 425,981 observations from 7,379. From the table it can be seen that the average page contains 5658 bytes of content and has undergone 84 revisions. However, the median is substantially lower at 3885 bytes and only 40 revisions. Also, the summary statistics of the first differences (variables starting with "Delta:") reveal that on a typical day nothing happens on a given page on Wikipedia. This highlights the necessity of using major events as a focal lense for analyzing activity on Wikipedia,[69] which is confirmed by the visual inspection of the direct and indirect effect of treatments.

[69]Further descriptive analyses that compare treated and control groups before and during treatment show that the groups are very similar in their activity levels before the shocks occurred and that the control group did not change its behavior during treatment. These tables and their description were omitted for reasons of brevity. They are available from the author upon request.

Table 16: Summary statistics: indirect neighbors of shocked articles (2 clicks away from epicenter) in the large events condition

	mean	sd	min	p10	p50	p90	max
Length of page (in bytes)	5,657	6,286	16	33	3,887	13,208	76,176
Number of authors	29	34	1	1	18	71	435
Clicks	33	174	0	0	0	70	29,865
Number of Revisions	84	133	1	2	40	211	2083
Links from Wikipedia	123	447	0	5	31	269	27,611
Dummy: literature section	.2	.4	0	0	0	1	1
Number of images	1.3	2.4	0	0	0	4	57
Number language links	13	18	0	0	7	37	179
References (footnotes)	1.3	4.2	0	0	0	4	150
Links to further info	2.7	5.1	0	0	1	7	130
time variable (normalized)	0	8.4	-14	-12	0	12	14
Delta: Number of Revisions	.035	.35	0	0	0	0	44
Delta: Length of page	1.8	106	-22,416	0	0	0	27,500
Delta: Number of authors	.013	.12	0	0	0	0	11
Delta: Links from Wikipedia	.049	2.5	-1,148	0	0	0	216
Delta: Number of images	.00047	.078	-27	0	0	0	20
Delta: References	.0014	.13	-32	0	0	0	29
Delta: Links further info	.0011	.12	-15	0	0	0	31

NOTES: The table shows the distribution of the main variables. The unit of observations is the outcome of a page i on day t. The time variable is normalized and runs from -14 to 14.; no. of obs. = 426,213; no. of start pages = 46; no. of articles = 7,383.

In Figure 6 I plot the average clicks (top row) and the average number of added revisions (bottom) for the three groups of pages zero clicks away (left column), one click away (middle column) and two clicks away (right column). Each of the plots features four lines. The bold blue line represents the treated group or its neighbors when they were actually treated, hence “treated in treatment phase”. The dashed red line represents the same group but during the placebo treatment at an earlier point in time. The thin green line (“not treated, treatment phase”) shows the control group at the time when the real shock occurred and the thin dotted yellow line represents the “unrelated” observations, which are never treated and observed in the placebo period.[70] The left column shows the control group and the article about the incident (“event page”; L_0), which are created only after the onset of the event. Most of these 23 pages did not exist at all before the onset of the event and therefore only a few have a placebo condition available. The upper row shows that the directly affected pages experience a very large spike of 8,500 clicks per day on average. The number of additional revisions peaks on the first days of treatment, when the pages are created: an average of almost forty revisions are added to a page on the first day. On the pages that are to share a reciprocal a link from the treated page the effect is quite large: while the clicks on the average L_1 page increase by 2,500, the absolute value of the average increase in revision activity is only five. When I look at pages that are two clicks away, the effects are much smaller, especially for revisions, but quite pronounced. The clicks on the average adjacent page go up by 35 and the absolute value of the average increase in revision activity is already no more than 0.04.

D.1.3 Disasters - Results

Before I proceed with the details of my estimations, it is worth recalling a few important facts. For large events the focus is on estimating Equation 12, for nodes that are two clicks away from the new disaster page. (cf. the equations in Section 3.4). This is due to two reasons: first, large events differ from “Today’s featured articles” in how local the treatments are. Second, while “Today’s featured articles” exist at the time of treatment, the page at the center of a large event treatment typically does not exist and will instead be created in the following days.

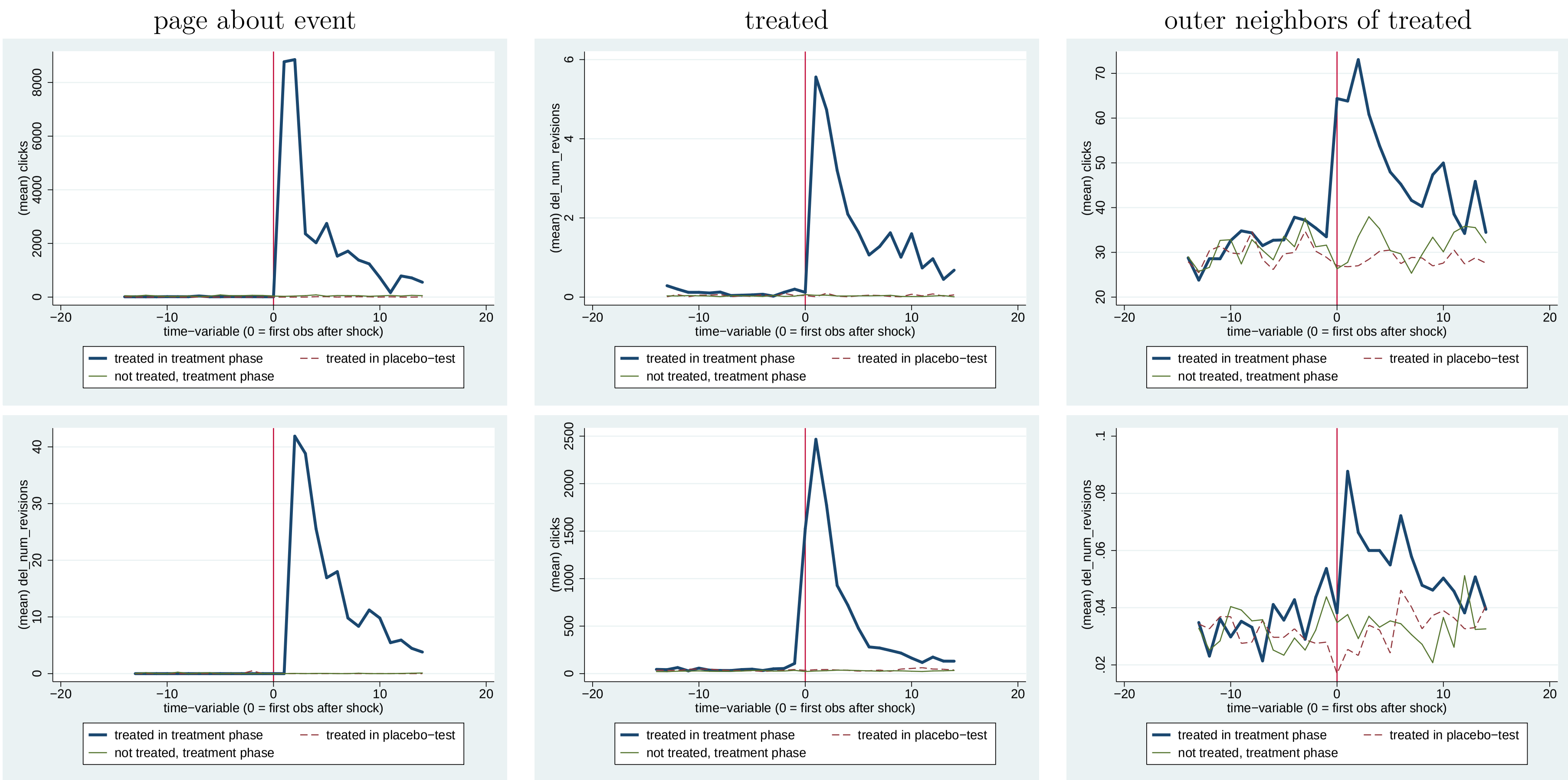
[70] For greater ease of representation I included a graphical representation of only two variables. The summary statistics for these groups before and after treatment are also available as tables upon request.

Table 17: Included disasters, associated indirect neighbors and observations.

name of event	pages No.	observations control	treated	Total
Air-France-Flug_447 [Air France Flight 447]	102.0	4,495.0	1,392.0	5,887.0
Air-India-Express-Flug_812 [Air India Express Flight 812]	369.0	19,662.0	1,711.0	21,373.0
Amoklauf_von_Winnenden [Winnenden school shooting]	74.0	2,088.0	2,146.0	4,234.0
Bahnunfall_von_Halle_(Belgien) [Halle train collision]	52.0	2,436.0	580.0	3,016.0
British-Airways-Flug_38 [British Airways Flight 38]	144.0	6,699.0	1,624.0	8,323.0
Buschfeuer_in_Victoria_2009 [2009 Victoria bush fire]	33.0	928.0	957.0	1,885.0
Deepwater_Horizon	203.0	8,178.0	3,596.0	11,774.0
Erdbeben_in_Haiti_2010 [2010 Haiti earthquake]	379.0	15,602.0	6,322.0	21,924.0
Erdbeben_in_Sichuan_2008 [2008 Sichuan earthquake]	227.0	11,571.0	1,508.0	13,079.0
Erdbeben_von_L'Aquila_2009 [2009 L'Aquila earthquake]	96.0	3,654.0	1,885.0	5,539.0
Flugzeugabsturz_bei_Smolensk [Smolensk plane crash]	368.0	12,412.0	8,758.0	21,170.0
Grubenunglück_von_San_Jose [San Jose mine disaster]	149.0	8,033.0	551.0	8,584.0
Josef_Fritzl	129.0	6,264.0	1,044.0	7,308.0
Kaukasuskrieg_2008 [Caucasian war 2008]	346.0	18,705.0	1,276.0	19,981.0
Kolontár-Dammbruch [Kolontar dam failure]	99.0	4,669.0	1,073.0	5,742.0
Luftangriff_bei_Kunduz [Kunduz airstrike]	2,107.0	113,767.0	7,772.0	121,539.0
Northwest-Airlines-Flug_253 [Northwest Airlines Flight 253]	1,151.0	65,279.0	1,276.0	66,555.0
Sumatra-Erdbeben_vom_September_2009 [2009 Sumatra earthquakes]	116.0	4,002.0	2,726.0	6,728.0
US-Airways-Flug_1549 [US Airways Flight 1549]	226.0	7,888.0	5,220.0	13,108.0
Unglück_bei_der_Loveparade_2010 [Love parade disaster (2010)]	499.0	15,283.0	13,572.0	28,855.0
Versuchter_Anschlag_am_Times_Square [2010 Times Square car bombing attempt]	202.0	10,353.0	1,334.0	11,687.0
Wald-_und_Torfbrände_in_Russland_2010 [2010 Russian wildfires]	273.0	13,485.0	2,204.0	15,689.0
Zugunglück_von_Castelldefels [Castelldefels train accident]	35.0	1,508.0	493.0	2,001.0
Total	7,379.0	356,961.0	69,020.0	425,981.0

NOTES: The table shows the events in the dataset. Column 1 shows the number of pages that are two clicks away from one of the two associated start pages (be it treated or control). Columns 2-4 show the number observations associated with the articles. Observations associated with actually treated articles are shown separately from control observations. Pages can be accessed by pasting the title behind the last slash in: *http://de.wikipedia.org/wiki/*

Figure 6: Catastrophes: Comparing average clicks (new edits) of treated pages (and neighbors) to three comparison groups.



Notes: The figure shows the results for natural disasters and large accidents. The left column shows the average effect on the pages about the disaster ("event pages" - by definition, they were created after the event), the middle column the directly treated pages, that users turn to, until the event gets a page of its own ("L1", with reciprocal link to the future event page), and the right column for the pages that are one click away from L_1. The upper row shows the average number of clicks the lower row shows the average number of edits. The outcomes are shown for the treated articles and the control groups separately. Directly hit pages received up to 8,500 additional clicks and up to 40 new revisions on average. Pages that will have a reciprocal link received up to approx. 2,500 clicks and up to 5 additional revisions. However, not only the treated pages, but also their neighbors received 35 additional clicks and up to 0.04 additional revisions on average.

$$y_{it} = \phi_i^{L_2} + \sum_{s \in S} \phi_{1,s}^{L_2} \lambda_s + \sum_{s \in S} \phi_{2,s}^{L_2} (\lambda_s * treat_{L_2,i}) + \xi_{it} \tag{12}$$

Hence, I present the results for the set of L_2 pages that are two clicks away from the epicenter: the future page about the disaster. Closer pages are no less interesting, but the shock of the analyzed events is very big and likely to directly affect pages that will eventually be directly and bidirectionally linked. If, for example, a city in the province under consideration was hit by the earthquake, the added content on that page might simply cover this very fact. In such a case, this is not an improvement that arose from the increased attention that results from the adjacent event, but a change that is directly caused by the treatment. As explained above, this is not the effect I am primarily interested in. Consequently I focus on pages that were indirectly linked at the time of the shock and that never became directly linked. These articles are no longer likely to be directly affected by the treatment on the page two clicks away.[71]

Otherwise I proceed as for "Today's featured articles" to avoid potentially endogenous link formation and exclude implicitly treated articles and "link-hubs." Moreover, to ensure that my L_2 pages are not directly related to the event I checked whether a page that was in L_2 when I evaluated the network a week before the shock was going to be linked to the page of the disaster later. Since this indicates a potential direct relationship, I eliminated such pages from the sample.

The results of the estimation of the model for L_2 nodes are shown in Table 18. The table shows the results for clicks in the first three columns and the results for the number of added revisions in Columns 4,5 and 6. All the specifications are OLS panel regressions which include a fixed effect for the page and standard errors are clustered on the event level (23 clusters). Column (1) and (4) shows the results of a simple before and after. Columns 2, 3, 5 and 6 show the contrast in the difference-in-differences. Note that I run each regression twice to take advantage of my two comparison groups: first I contrast the treated pages against the control group and then I contrast it with the placebo treatment, i.e. with the treated articles themselves, but simulating a (placebo) treatment 42 days (i.e. 7 weeks) before the real shock.

The table shows the coefficients for the the period of the shock and the two subsequent days individually. Periods before the shock are represent in dummies that average over several days, and the periods later than two days after the shock are represented analogously. The reference group are days -14 to -8 before the advertisement. As in the visual evidence, the average increase in clicks relative to the control group (Column 2) amounts to 35.5-38.9 more clicks on average. Using the placebo treatment as comparison group (Column 3) this effect is almost equal, but a bit larger from the second day onwards.

Does the spillover in attention also translate to additional content generation? Obviously, this question matters for the relevance of the spillovers I find in this paper. If it does, spillovers of attention have far-reaching implications for other peer production settings.[72]

The effects are smaller for the number of revisions (in line with the graphical analysis). An effect is consistently revealed from the first day after the treatment. It is small in absolute terms, since roughly one in twenty-five pages gets an additional revision. Yet, given the low average activity, this is still a noteworthy effect. Comparing the pages with the placebo treatment I observe a small increase in editing activity before the onset of the event, which is not confirmed by comparison with the control group. The size of the effect still more than doubles after day 1 which suggests a drastic increase in editing activity.[73]

Finally it's possible to estimate the bounds on α like in Section 5.3. Evaluating at the median,

[71] The results for the L_1 group are included in the appendix. The effects are very large and statistically significant. The estimated coefficients for the L_0 group (not reported) are close to 4,500 for clicks and between 20 and 25 for revisions. However, due to the lack of sufficient observations, even these very large coefficient estimates are not statistically different from zero.

[72] If more attention leads to better or more contributions, the importance of link networks for channeling attention would have important implications for open source software, research activities and innovation.

[73] I verified that the result is not driven by running a robustness check, where I exclude four events: the event which was associated to most pages in my dataset, Tunisia and those where the starting date or the most important page of the event was most difficult to identify: the bancrupcy of Lehman, the eruption of Eyjafjallajykull and the plane crash in Smolensk. In this specification, the results are confirmed. The most notable difference is the increased magnitude of the effect in the clicks, as for the remaining events, the average increase is close to 15 additional clicks. Despite the fact, that there are still more than 6,000 pages included in both comparisons, the effects for the number of revisions are no longer significantly different from zero, except in the fourth period of one specification.

Table 18: Relationship of clicks/added revisions and time dummies for indirect neighbors of shocked articles (2 clicks away from epicenter) in the large events condition.

	clicks			del revisions		
	(1) before after	(2) compare control	(3) comp. placebo	(4) before after	(5) comp. control	(6) comp. placebo
Before: days - 7 to -4	3.493*	2.311	3.547	0.007	0.012**	0.009**
	(1.741)	(2.039)	(2.123)	(0.004)	(0.005)	(0.004)
Before: days - 3 to -1	5.155**	1.291	3.288	0.015**	0.010	0.023***
	(2.263)	(4.242)	(2.922)	(0.006)	(0.007)	(0.006)
t = 0	34.163**	37.479**	36.572**	0.057**	0.052**	0.065***
	(14.583)	(14.557)	(14.528)	(0.023)	(0.024)	(0.023)
t = 1	33.598***	35.476***	36.229***	0.036***	0.039***	0.045***
	(11.036)	(11.059)	(11.028)	(0.012)	(0.014)	(0.012)
t = 2	42.867***	38.949***	45.286***	0.029***	0.025**	0.028***
	(13.741)	(13.733)	(13.633)	(0.009)	(0.011)	(0.010)
After: days 3 to 6	21.758***	18.075***	22.010***	0.031**	0.030**	0.028**
	(5.336)	(5.713)	(5.433)	(0.012)	(0.012)	(0.012)
After: days 7 to 14	11.408**	15.705***	11.962**	0.017*	0.023**	0.018
	(4.847)	(5.267)	(5.114)	(0.009)	(0.010)	(0.011)
Time Dummies	No	Yes	Yes	No	Yes	Yes
Mean dep. Variable	41.053	34.108	35.183	0.046	0.037	0.038
Observations	52426	162426	104346	52426	162426	104346
Number of Pages	2383	7383	4743	2383	7383	4743
Adj. R^2	0.003	0.003	0.003	0.002	0.001	0.001

NOTES: The table shows the results of the reduced form regressions estimating the ITE . Columns (1)-(3) show the results for clicks and Columns (4-6) for new edits to the articles. Specification (1) and (4) show a simple 'before and after'; (2) and (5) contrast treated and comparison group; Columns (3) and (6) show the comparison of treated articles with themselves but seven weeks earlier (placebo treatment). Fixed Effects Panel-Regressions with heteroscedasticity robust standard errors. The unit of observations is the outcome of a page i on day t. The time variable is normalized and runs from -14 to 14.; Only crossterms on and shortly after the day of treatment are shown individually, all others are shown as summarized in groups. Reference group t-14 to t-8; standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; no. of obs. = 323,334; no. of clusters = 46; no. of articles = 7,383.

indirect neighbors of disaster pages have themselves 34 neighbors and I observe 2,440 additional clicks on directly affected pages. The upper bound is 0.483. The lower bound, $\underline{\hat{\alpha}}$ is 0.320. Since these conditions differ in the source of the shock and in the sets (L_1 and L_2), it is reassuring, that both the direct effects and even the estimated bounds are close in the two conditions.

D.2 Robustness Checks - Alternative Specifications

This section shows further robustness checks. I ran the estimation separately only for advertisements that occurred during weekdays. If a part of my effect were driven by user-idleness, the spillovers should be smaller during the working days. The results in Table 19 show no discernible difference. In the second check I investigate if my sampling on 3 days of the month is important. I further restrict the sample to advertisements that came from the 10^{th} day of each month (cf. Table 20). This reduces the number of advertisements to 34, but otherwise does not substantially change the main result. However, there seems to be a slight reduction in the activity on the days before advertisement for this subsample.

Table 19: Robustness Check: Relationship of clicks/added revisions and time dummies for direct neighbors of shocked articles in the 'featured articles' condition only for advertisements during weekdays.

	clicks		del revisions		del authors	
	(1) compare control	(2) comp. placebo	(3) comp. control	(4) comp. placebo	(5) comp. control	(6) comp. placebo
Before: days - 7 to -4	0.950	0.647	-0.001	0.002	0.000	0.000
	(1.267)	(1.189)	(0.004)	(0.004)	(0.002)	(0.002)
Before: days - 3 to -1	2.475	2.164	0.000	-0.001	-0.001	-0.002
	(2.794)	(2.207)	(0.005)	(0.005)	(0.002)	(0.002)
t = 0	29.677***	36.267***	0.026**	0.030***	0.012***	0.013***
	(7.028)	(6.760)	(0.010)	(0.009)	(0.004)	(0.004)
t = 1	2.081	2.216	0.003	0.011	0.002	0.001
	(2.124)	(2.350)	(0.007)	(0.008)	(0.002)	(0.003)
t = 2	-0.293	-1.448	-0.015*	-0.010	-0.004*	-0.004*
	(2.840)	(2.437)	(0.008)	(0.007)	(0.003)	(0.002)
After: days 3 to 6	-2.377	-0.571	-0.000	-0.002	-0.001	-0.001
	(2.136)	(2.511)	(0.005)	(0.005)	(0.002)	(0.002)
After: days 7 to 14	-1.536	0.784	-0.003	-0.003	-0.002	-0.002
	(2.980)	(2.975)	(0.008)	(0.008)	(0.003)	(0.002)
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	37.213	37.622	0.046	0.046	0.016	0.016
Observations	280568	311854	280568	311854	280568	311854
Number of Pages	15732	16881	15732	16881	15732	16881
Adj. R^2	0.003	0.003	0.000	0.000	0.000	0.000

Standard errors in parentheses.
Fixed Effects Panel-Regressions with heteroscedasticity robust standard errors.
Only crossterms closer to treatment are shown, but all were included. Reference group t-14 to t-8.
* p<0.10, ** p<0.05, *** p<0.01

Table 20: Robustness Check: Relationship of clicks/added revisions and time dummies for direct neighbors of shocked articles in the 'featured articles' condition only for advertisements the 10th of each month.

	clicks		del revisions		del authors	
	(1) compare control	(2) comp. placebo	(3) comp. control	(4) comp. placebo	(5) comp. control	(6) comp. placebo
Before: days - 7 to -4	0.815	0.379	0.006	0.006	-0.000	-0.000
	(1.835)	(1.745)	(0.008)	(0.007)	(0.002)	(0.002)
Before: days - 3 to -1	-1.258	0.866	-0.016*	-0.004	-0.008**	-0.005**
	(3.181)	(3.438)	(0.008)	(0.007)	(0.004)	(0.002)
t = 0	33.932***	34.713***	0.034**	0.035**	0.013***	0.013***
	(9.237)	(9.387)	(0.013)	(0.014)	(0.005)	(0.005)
t = 1	0.646	0.848	0.017*	0.019	0.004	0.002
	(2.544)	(3.524)	(0.010)	(0.011)	(0.004)	(0.004)
t = 2	-1.719	-3.625	-0.009	-0.010	-0.006	-0.006*
	(3.101)	(3.261)	(0.012)	(0.013)	(0.005)	(0.003)
After: days 3 to 6	-3.821	-1.620	-0.002	-0.003	-0.002	-0.003
	(2.411)	(3.666)	(0.007)	(0.008)	(0.002)	(0.002)
After: days 7 to 14	0.900	1.518	-0.007	-0.010	-0.004	-0.004
	(4.011)	(4.299)	(0.012)	(0.011)	(0.003)	(0.003)
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. Variable	32.722	35.688	0.045	0.047	0.016	0.016
Observations	120758	166518	120758	166518	120758	166518
Number of Pages	5489	7569	5489	7569	5489	7569
Adj. R^2	0.004	0.003	0.000	0.000	0.000	0.000

Standard errors in parentheses.
Fixed Effects Panel-Regressions with heteroscedasticity robust standard errors.
Only crossterms closer to treatment are shown, but all were included. Reference group t-14 to t-8.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

E The Empirical Model and Structural Identification of the Parameter of Interest.

This section presents the structural model and discusses the parameters of interest, the challenges in identifying them and the approach taken to tackle them.

E.1 Introductory Remarks

I depart from the well known linear-in-means model as formulated by Manski (1993).[74]

$$y_{it} = \alpha \frac{\sum_{j \in P_{it}} y_{jt}}{N_{P_{it}}} + X_{it-1}\beta + \gamma \frac{\sum_{j \in P_{it}} X_{jt-1}}{N_{P_{it}}} + \epsilon_{it} \tag{13}$$

y_{it} denotes the outcome of interest in period t and X_{it-1} are i's observed characteristics at the end of period $t-1$ (beginning of period t).[75] P_{it} is the set of i's peers and $N_{P_{it}}$ represents the number of i's peers. β measures the effect of i's own characteristics and γ accounts for how i's performance is affected by the peers' average characteristics. The coefficient of interest is α. In the present context it measures how the clicks on page i are influenced the clicks on the adjacent pages. Bramoullé et al. (2009) suggest a more succinct notation based on vector and matrix notation:

$$\mathbf{y_t} = \alpha \mathbf{G} \mathbf{y_t} + \beta \mathbf{X_{t-1}} + \gamma \mathbf{G} \mathbf{X_{t-1}} + \epsilon_\mathbf{t} \quad \mathbf{E}[\epsilon_\mathbf{t} | \mathbf{X_{t-1}}] = \mathbf{0}$$

Note that the linear in means model provides the weakest basis for identification. I conjecture that the insights carry over to other linear models and less weakly identified non-linear models.

E.2 Setup and Basic Intuition

Augment the model (eq. 13) by observable and locally applied treatments (shocks):

$$y_{it} = \alpha \frac{\sum_{j \in P_{it}} y_{jt}}{N_{P_{it}}} + X_{i,t-1}\beta + \gamma \frac{\sum_{j \in P_{it}} X_{j,t-1}}{N_{P_{it}}} + \delta_1 D_{it} + \epsilon_{it} \tag{14}$$

where the new coefficient δ_1 measures the direct effect if a node(page) is treated.

Note that $X_{it-1}\beta$ may contain an individual fixed effect and an additively separable age-dependent part: $X_{it-1}\beta = \beta_i + \widetilde{X_{i,t-1}}\beta_1 + \beta_2 f(age)$. To see how local treatments can be used as a source of identification, consider two pairs of nodes.

Local application of treatment: First, consider two connected nodes, where one is treated ($\ell 0$) in period t and the neighbors are not treated ($\ell 1 \in L1$). Assume for simplicity that $\ell 0$ is the only treated node in $\ell 1$'s neighborhood.

$$\ell 0 :: \boldsymbol{y_{\ell 0 t}} = \alpha \frac{\sum_{j \in P_{\ell 0 t}} y_{jt}}{N_{P_{\ell 0 t}}} + X_{\ell 0 t-1}\beta + \gamma \frac{\sum_{j \in P_{\ell 0 t}} X_{jt-1}}{N_{P_{\ell 0 t}}} + \delta_1 \boldsymbol{1} + \epsilon_{\ell 0 t} \tag{15}$$

$$\ell 1 \in L1 :: y_{\ell 1 t} = \alpha \frac{\boldsymbol{y_{\ell 0 t}} + \sum_{j \in P_{\ell 1 t} / \ell 0} y_{jt}}{N_{P_{\ell 1 t}}} + X_{\ell 1 t-1}\beta + \gamma \frac{\sum_{j \in P_{\ell 1 t}} X_{jt-1}}{N_{P_{\ell 1 t}}} + \delta_1 0 + \epsilon_{\ell 1 t} \tag{16}$$

If we now consider a comparison group of two connected nodes ($c0$ and $c1$) where nobody gets treated, D_t would take the value 0 for both $c0$ and $c1$. The newly introduced term would simply

[74]Note that it is easy to add a fixed effect to the model, but that it will be eliminated when taking differences. Consequently, I omit it for ease of notation.

[75]The choice of the temporal structure depends on the application that the researcher has in mind. In the present application many independent variables are stock variables (articles' characteristics such as page length), while the dependent variables are typically flows (clicks or new revisions).

drop out. It can easily be seen, how the local treatment will allow to measure the spillover or peer effect. This will be possible despite the richness in other sources of variation, provided (i) the shocks are large enough and (ii) the "control network" allows to credibly infer the dynamics in the "treated network", had no treatment taken place.

Condensed Notation: I use the matrix notation suggested by Bramoullé et al. (2009) and incorporate the newly proposed vector of treatments[76]:

$$\mathbf{y_t} = \alpha\mathbf{G}\mathbf{y_t} + \mathbf{X_{t-1}}\beta + \gamma\mathbf{G}\mathbf{X_{t-1}} + \delta_1\mathbf{D_t} + \epsilon_\mathbf{t} \qquad \mathbf{E}[\epsilon_\mathbf{t}|\mathbf{D_t}] = \mathbf{0} \tag{17}$$

$\mathbf{G}$ is a NxN matrix, which captures the link structure in the network. $G_{ij} = \frac{1}{N_{P_i}-1}$ if i receives a link from j and $G_{ij} = 0$ otherwise. Note that I do not require $\mathbf{G}$ to be exogenously given, but only $\mathbf{D_t}$, a vector which is 1 at the treated nodes (if they are *currently* treated) and 0 otherwise. In some of the proofs and in my application I will assume a local treatment that affects only a single node. Formally this is written as an elemenatry vector $\mathbf{D_t} = \mathbf{e_{\ell 0}}$ with the 1 in the coordinate that corresponds to the treated node. On the untreated subnetwork we have $\mathbf{D_t} = \mathbf{0}$, a vector of zeros.

Unlike Bramoullé et al. (2009), I do not look for an instrument for $\mathbf{Gy}$. Since I rather use exogenous shocks that affect only one part of the network, there will be no requirements on the linear independence of $\mathbf{G}$ and $\mathbf{G^2}$.

E.3 Proof of Result 1

The assumptions are as stated in the main text before the result. I allow $\mathbf{X_t}$ to change over time and consider the S-period difference-in-differences.[77]

Proof. The reduced form corresponding to equation 17 is given by:

$$\mathbf{y_t} = (\mathbf{I} - \alpha\mathbf{G})^{-1}[\mathbf{X_{t-1}}\beta + \gamma\mathbf{G}\mathbf{X_{t-1}} + \delta_1\mathbf{D_t} + \epsilon_\mathbf{t}] \tag{18}$$

and the expectation conditional on the "treatment" is:

$$\begin{aligned} \mathbf{E}[\mathbf{y_t}|\mathbf{D_t}] &= (\mathbf{I} - \alpha\mathbf{G})^{-1}[(\beta + \gamma\mathbf{G})\mathbf{E}[\mathbf{X_{t-1}}|\mathbf{D_t}] + \delta_1\mathbf{D_t} + \mathbf{E}[\epsilon_\mathbf{t}|\mathbf{D_t}]] =^{b.A.} \\ &=^{b.A.} (\mathbf{I} - \alpha\mathbf{G})^{-1}[(\beta + \gamma\mathbf{G})\mathbf{E}[\mathbf{X_{t-1}}|\mathbf{D_t}] + \delta_1\mathbf{D_t}] \end{aligned} \tag{19}$$

Taking the first difference, we obtain:

$$\begin{aligned} \mathbf{\Delta_t}\mathbf{E}[\mathbf{y}|\mathbf{D}] = \; & \mathbf{E}[\mathbf{y_t}|\mathbf{D_t}] - \mathbf{E}[\mathbf{y_{t-1}}|\mathbf{D_{t-1}}] = \\ = \; & (\mathbf{I} - \alpha\mathbf{G})^{-1}[(\beta + \gamma\mathbf{G})\{\mathbf{E}[\mathbf{X_{t-1}}|\mathbf{D_t}] - \mathbf{E}[\mathbf{X_{t-2}}|\mathbf{D_{t-1}}]\} + \delta_1\mathbf{\Delta D_t}] = \\ = \; & (\mathbf{I} - \alpha\mathbf{G})^{-1}[(\beta + \gamma\mathbf{G})\{\mathbf{E}[\mathbf{X_{t-1}}|\mathbf{D_t}] - \mathbf{E}[\mathbf{X_{t-2}}|\mathbf{D_{t-1}}]\} + \delta_1\mathbf{D_t}] \end{aligned} \tag{20}$$

...where $\mathbf{\Delta D_t} = \mathbf{D_t} - \mathbf{D_{t-1}}$ and the second equality holds, because treatments are assumed to start in period t, but not before.[78]

Now consider the control group formed by the same network, but S periods earlier:

$$\mathbf{y_{t-S}} = \alpha\mathbf{G}\mathbf{y_{t-S}} + \mathbf{X_{t-S-1}}\beta + \gamma\mathbf{G}\mathbf{X_{t-S-1}} + \delta_1\mathbf{D_{t-S}} + \epsilon_\mathbf{t-S}$$

[76] $\mathbf{X}$ might include a time-dependent component (e.g. a linear function of age) as well.

[77] The S-period difference is reasonably close to the "placebo condition" of my application below. At the end of the formal derivations I will discuss the consequences of relaxing the requirement of a stable network or the consequences of adding the assumption that $\mathbf{X_t}$ does not change between the periods of observation. Importantly the nodes in the network have to be observed over time and have to evolve in a stable fashion, to ensure that the first differences are the same at t and $t-S$. The setting also corresponds to comparing the evolution of nodes in a very stable network during a post and a pre-treatment stage.

[78] That difference contains the time-dependent component and the effect of any changes in the independent variables. If βX_{it} is modeled to contain an additively separable age-dependent part as in our example above, $\Delta X_{it-S}\beta$ would contain $\frac{df(age)}{dt}$ (to be eliminated by taking the Difference in Differences).

The first difference of the reduced form's conditional expectations are:

$$
\begin{aligned}
\mathbf{\Delta_{t-S}E[y|D]} = \quad & \mathbf{E[y_{t-S}|D_{t-S}]} - \mathbf{E[y_{t-S-1}|D_{t-S-1}]} = \\
& = (\mathbf{I} - \alpha\mathbf{G})^{-1}[(\beta + \gamma\mathbf{G})\{\mathbf{E[X_{t-S-1}|D_{t-S}]} - \mathbf{E[X_{t-S-2}|D_{t-S-1}]}\} + \delta_1\mathbf{\Delta D_{t-S}}] = \\
& = (\mathbf{I} - \alpha\mathbf{G})^{-1}[(\beta + \gamma\mathbf{G})\{\mathbf{E[X_{t-S-1}|D_{t-S}]} - \mathbf{E[X_{t-S-2}|D_{t-S-1}]}\} + 0]
\end{aligned}
$$

with $\mathbf{\Delta D_{t-S}} = 0$, since treatments are assumed to start in period t, but not earlier. Proceeding to take the Difference in Differences, we obtain:

$$
\begin{aligned}
\mathbf{DiD} := \quad & \mathbf{\Delta y_t E[y|D]} - \quad \mathbf{\Delta y_{t-S}E[y|D]} = \\
= \quad & (\mathbf{I} - \alpha\mathbf{G})^{-1} \quad [(\beta + \gamma\mathbf{G})\{\mathbf{E[X_{t-1}|D_t]} - \mathbf{E[X_{t-2}|D_{t-1}]}\} + \delta_1\mathbf{D_t}] - \\
& \quad - \quad (\beta + \gamma\mathbf{G})\{\mathbf{E[X_{t-S-1}|D_{t-S}]} - \mathbf{E[X_{t-S-2}|D_{t-S-1}]}\}]
\end{aligned}
$$

Denoting the change in the expectation of $\mathbf{X_{t-1}}$ conditional on $\mathbf{D_t}$ more concisely by $\{\mathbf{E[X_{t-1}|D_t]} - \mathbf{E[X_{t-2}|D_{t-1}]}\} = \mathbf{\Delta_t(E[X|D])}$ and rearranging gives:

$$
\mathbf{DiD} = (\mathbf{I} - \alpha\mathbf{G})^{-1} \quad [(\beta + \gamma G)\{\mathbf{\Delta_t(E[X|D])} - \mathbf{\Delta_{t-S}(E[X|D])}\} + \delta_1\mathbf{D_t}] \tag{21}
$$

which reduces to:

$$
\mathbf{DiD} = (\mathbf{I} - \alpha\mathbf{G})^{-1}\{\delta_1\mathbf{D_t}\} \tag{22}
$$

if $\mathbf{\Delta_t(E[X|D])} = \mathbf{\Delta_{t-S}(E[X|D])}$. Thus, the identifying assumption is that the expected changes of the pages between $t-1$ and t are the same as from $t-S-1$ and $t-S$. This is satisfied if $\Delta X_t|D_t$ is stationary of order one.

Provided $(\mathbf{I} - \alpha\mathbf{G})^{-1}$ is invertible we can use the property that $(\mathbf{I} - \alpha\mathbf{G})^{-1} = \sum_{s=0}^{\infty} \alpha^s G^s$[79], the general impact of a local treatment is:

$$
\mathbf{DiD'} \quad = \quad \delta_1\mathbf{D_t'}(\mathbf{I} + \alpha\mathbf{G} + \alpha^{\mathbf{2}}\mathbf{G^2} + \alpha^{\mathbf{3}}\mathbf{G^3} + ...)' \tag{23}
$$

which completes the proof. ■

Discussion of the assumptions used:

1. $\mathbf{E}[\epsilon_\mathbf{t}|\mathbf{D_t}] = \mathbf{0}$ - The shock is independent of the error term.

2. α is smaller than the norm of the inverse of the largest eigenvalue of $\mathbf{G}$. A regularity condition to ensures that the expression $(\mathbf{I} - \alpha\mathbf{G})^{-1} = \sum_{s=0}^{\infty} \alpha^s G^s$ is well defined.

3. I assumed the network to be stable over time and used it's earlier state as control observation. Formally this is written as $\mathbf{G_{\ell,t}} = \mathbf{G_{\ell,t-1}} = \mathbf{G}$ and $\mathbf{G_{c,t}} = \mathbf{G_{\ell,t-S}} = \mathbf{G}$. This assumption could be relaxed, but only at the expense of strengthening the following assumption.

4. $\mathbf{\Delta_t(E[X|D])} - \mathbf{\Delta_{t-S}(E[X|D])}$, which means that the expected changes of the pages between $t-1$ and t are the same as from $t-S-1$ and $t-S$[80]. This is the analogue of the well known common trends assumption.

5. SUTVA on the level of subnetworks: the non-treated subnetwork is not affected by treatment of the treated subnetwork. In the present context SUTVA holds for my placebo condition and, given the size of the Wikipedia network, it is also plausibly satisfied for the control group formed by a remote part of the network.

The proof for the control group consisting of remote nodes is analogous. It relaxes the third assumption and requires a more general formulation of the fourth. The qualitative meaning of

[79] $(\mathbf{I} - \alpha\mathbf{G})$ is invertible if $\alpha < 1$ (Bramoullé et al. (2009)) and the infinite sum is well defined if α is smaller than the norm of the inverse of the largest eigenvalue of $\mathbf{G}$ (Ballester et al. (2006)).

[80] Particularly, any time trends or other dynamics, is to be eliminated by the Differences in Differences, if $\frac{df(age)}{dt}$ is the same evaluated at t-S and at t.

the generalized assumption will be the same: Absent treatment the treated *network* and the control *network* must "*evolve* in the same way."[81] However, I have to maintain the assumption that the network formation *process* is not affected by the treatment.[82] I do not consider this assumption warranted for disasters and I checked this assumption in my "today's featured article application": Link formation remains on low levels. On normal days, articles' degree grows steadily by about 0.1 links per day, with total in-links averaging at 120. There is a short increase by 0.2 in-links per article (or 0.2% of the link stock). Yet, first this is in sync with the peak in edits, but not with the peak in clicks, and second, like for edits, the peak is large in relative, but small in absolute terms. I conclude that this is an acceptably small source of potential bias.

E.3.1 Estimating α: Analysis on the Node Level

Above we have shown what is measured by the difference-in-differences. From now on I shall refer to a node in the control condition by c and to a node in the treated condition by ℓ. If $\mathbf{D_t}$ denotes the vector of treatments which is 1 at the treated nodes and 0 otherwise, estimation of the difference-in-differences identifies:

$$\mathbf{DiD}' \quad = \quad \delta_1 \mathbf{D_t'}(\mathbf{I} + \alpha\mathbf{G} + \alpha^{\mathbf{2}}\mathbf{G^2} + \alpha^{\mathbf{3}}\mathbf{G^3} + ...)' \tag{24}$$

When we take the analysis back from the level of treated networks and look at the nodes individually, all that matters for each focal node j is its own row in this set of equations. To simplify this analysis I now introduce the local treatment assumption, exploiting the fact that only a single node in my network is treated each day. This is like a partial population treatment Moffitt (2001) with only one single node (a mini population) being treated.

Local Treatment Assumption: *Under the local treatment assumption* $\mathbf{D_t} = \mathbf{e_i}$*, where* $\mathbf{e_i}$ *is an elementary vector with node* i *being the only treated node.*

If only one node is treated, the spillover dynamic is greatly simplified. With $\mathbf{D} = \mathbf{e_i}$, the only factor to be evaluated for each node is its corresponding ji element in the matrix $\mathbf{G}$, $\mathbf{G^2}$ and its higher orders.[83] We distinguish a shocked node $\ell 0 \in L0$, a neighbor $\ell 1 \in L1$ and the indirect neighbors (2 clicks away, 3 clicks away etc.) as follows:

$$\begin{aligned} \ell 0: \; DiD_i &= \delta_1(1 + 0 + \alpha^2 G_{ii}^2 + \alpha^3 G_{ii}^3 + ...) \\ \ell 1: \; DiD_{j\in L_1} &= \delta_1(0 + \alpha G_{ij} + \alpha^2 G_{ij}^2 + \alpha^3 G_{ij}^3 + ...) \\ \ell 2: \; DiD_{k\in L_2} &= \delta_1(0 + 0 + \alpha^2 G_{ik}^2 + \alpha^3 G_{ik}^3 + ...) \\ &\qquad\qquad\qquad\qquad\qquad\qquad \textit{etc.} \end{aligned} \tag{25}$$

Sorting the nodes with respect to their distance from $\ell 0$ and estimating these strata separately results in as many estimation equations as can reasonably be traced and two parameters to be estimated. This fact is the basic idea of this paper, because it enables the researcher to back out the estimates for the structural parameters α and δ_1. All that is needed is a sequence of reduced form difference-in-differences estimates for increasingly large link distances. If the precise information on $\mathbf{G}$ and its higher orders is available the parameters can be directly estimated.[84]

[81] To be more precise, the link formation and the way in which the characteristics of the nodes change over time have to be the same (common trends) in both networks. This guarantees that the counterfactual outcome of the treated network can be inferred from its own past and the evolution in the control network. The derivations require a lot of notational overhead and the resulting conditions are quite unwieldy. Assumption 4 would refer not only to $\mathbf{\Delta X}$, but to $\mathbf{\Delta GX}$ to allow for relaxing Assumption 3.

[82] If this is the case, all estimates of indirect treatment effects, will reflect a sum of the treatment on the existing network and new spillovers due to the changes in the link network (cf. Comola and Prina (2013)), which will lead to upward biases if not accounted for.

[83] The information in the higher orders of the adjacency matrix $\mathbf{G}$ is the same as the information from the sampling strategy in combination with knowing who was affected by the local treatment. Some nodes (L0) are known to be directly treated. Neighbors (L1) have a direct link so that the entry in $\mathbf{G}$ that links them to the treated node is positive. However, for those who only have an indirect link, the corresponding entry in $\mathbf{G}$ takes the value 0 and only the relevant element of $\mathbf{G^2}$ will be greater than 0.

[84] To do this use all the ij values that correspond to each individual focal node j as weights for α, α^2,

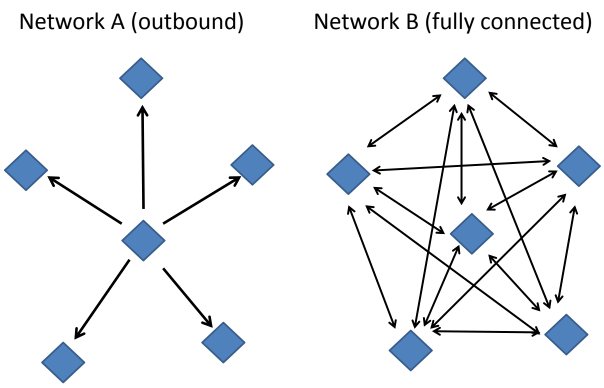
If not, it is possible to compute an upper and a lower bound for the parameters α and δ_1. In the next subsection I proceed to show how the boundary estimates can be computed.

E.4 Estimating Bounds for the Parameters of Interest

If the researcher lacks information on $\mathbf{G}$ it is possible to compute an upper and a lower bound for the social parameter α and the treatment effect δ_1. The goal in this section is to back out a lower and an upper bound estimate for α and δ_1, that is based only on the estimated DiD's and the number of nodes. This is useful, since the precise information on $\mathbf{G}$ is often not easy to obtain.[85] In my proofs I use the local treatment assumption (only one individual in the network is treated), for both ease of notation and understanding. It applies to "today's featured articles".[86]

In what follows I will show how to obtain these bounds. In Subsection E.4.1, I will give an intuitive account of the underlying ideas. In Subsection E.4.2, I will set up the preliminaries, including a Lemma that will be used. Subsection E.4.3 obtains the upper bound and Subsection E.4.4, finally, provides the proof for the lower bound.

Figure 7: Schematic representation of the two extreme networks, used to compute the upper and lower bound estimates of the parameters of interest.



NOTES: The " outbound network" (left) is used to obtain the upper bound estimate. It is a directed network with only "outward bound" links. Holding the number of nodes and the observed ITEs fixed, the social parameter will be estimated to be largest in this type of network. The fully connected network (right), is the benchmark case from which the lower bound of the social parameter can be estimated.

E.4.1 Intuition for Bounds

To see why we can bound the parameter, even without knowing the details of the network structure, we can select two 'specific 'extreme" types of networks which either minimize or maximize the higher order effects. For greater convenience, I repeat the illustration of such networks in Figure 7.

The network that minimizes higher order spillovers is a directed network with only "outward bound" links from $\ell 0$ to $\ell 1 \in L1$ [87]. This implies no links between the nodes in $L1$ and will

α^3, etc. and minimize a quadratic loss function. Unfortunately I cannot show this here, because the full matrix $\mathbf{G}$ formed by the German Wikipedia is too large to be computed in memory.

[85] The information might either not be available, or so big that computing its higher orders might confront the researcher with substantial computational challenges.

[86] I conjecture that extending the proof to partial population or randomized treatments will be straight forward. It merely means taking into account that more than one node gets treated and that the effects from the treated can also spill to the other treated, which will render the formulas quite unwieldy.

[87] and possibly further on to $\ell 2 \in L2$, $\ell 3 \in L3$ and so on.

serve as upper bound. The opposite type of network is a network, where every node is the direct neighbor of every one of its peers.[88] The fully connected network simplifies the analysis, because it has only two types of nodes (treated or not). Higher order spillovers are the same for every node of the same type. Moreover, given α and N, the fully connected network has the greatest second and higher order spillovers.[89] This allows to derive a closed form solution for the lower bounds of the relevant parameters.

E.4.2 Preliminaries

Before I proceed to characterize the bounds of the coefficient, it is useful to point out a fact that will be important in the argument that follows. Start by rewriting the formulas in equation 25 without explicit characterization of the higher order spills:

$$DiD_0 = \delta_1 + HO_{\ell 0} \tag{26}$$

$$DiD_1 = \frac{\alpha}{NP_{\ell 1}}\delta_1 + HO_{\ell 1} \tag{27}$$

where $HO_{\ell 0} = \delta_1(\alpha^2 G_{ii}^2 + \alpha^3 G_{ii}^3 + ...)$ and $HO_{\ell 1} = \delta_1(\alpha^2 G_{ij}^2 + \alpha^3 G_{ij}^3 + ...)$. These effects are typically not trivial. They depend on the underlying network of peers and need to take into account the network structure. However, I can use a simple insight concerning the size of the higher order effects.

Lemma 1 *Given the total effect, larger higher order effects, imply smaller coefficients, i.e. for* $DiD_0 > DiD_1 > HO^B > HO^A \geq 0$*: for any* $HO^A < HO^B$, $\alpha^A > \alpha^B$ *and* $\delta_1^A > \delta_1^B$.[90]

Proof. We have to make the following two comparisons:

$$DiD_0 = \delta_1^A + HO^A \quad vs. \quad DiD_0 = \delta_1^B + HO^B$$
$$DiD_1 = \frac{\alpha^A}{NP_{\ell 1}}\delta_1^A + HO^A \quad vs. \quad DiD_1 = \frac{\alpha^B}{NP_{\ell 1}}\delta_1^B + HO^B$$

This can be transformed as follows:

$$\delta_1^A = DiD_0 - HO^A \quad vs. \quad \delta_1^B = DiD_0 - HO^B \tag{28}$$

$$\alpha^A = \frac{(DiD_1 - HO^A)}{\delta_1^A}NP_{\ell 1} \quad vs. \quad \alpha^B = \frac{(DiD_1 - HO^B)}{\delta_1^B}NP_{\ell 1} \tag{29}$$

From equation 28 it is immediately obvious that $HO^A < HO^B$ implies $\delta_1^A > \delta_1^B$. For comparing α substitute the corresponding δ_1 from 28 into 29, define $HO^A := HO^B - \varepsilon$ (for $\varepsilon > 0$) and rewrite equation 29 as

$$\alpha^A = \frac{a}{b}NP_{\ell 1} \quad vs. \quad \alpha^B = \frac{a - \varepsilon}{b - \varepsilon}NP_{\ell 1} \tag{30}$$

where $a = (DiD_1 - HO^A)$ and $b = DiD_0 - HO^A$. Comparing α^A $vs.$ α^B is equivalent to

[88] I will sometimes refer to this network as "classroom" network.

[89] Every node affects every other node via a direct link and everybody will get second and higher round spillovers from *every* other node.

[90] Note that the requirement $DiD_1 > HO^B$ has bite, since it implies $\alpha < 0.5$. This assumption need not be satisfied in all applications, but it applies well to settings where the spills dissipate quickly and to settings where the direct effect on the treated is much larger than on the neighbors ($DiD_0 >> DiD_1$). This is the case in most applications and certainly so in the present one.

comparing $\frac{a}{b}$ $vs.$ $\frac{a-\varepsilon}{b-\varepsilon}$. Since we have $a, b, \varepsilon > 0$, $\varepsilon < b$ and $\varepsilon < a$:

$$\begin{aligned} \frac{a}{b} - \frac{a-\varepsilon}{b-\varepsilon} > 0 \quad &\Leftrightarrow \quad a(b-\varepsilon) - b(a-\varepsilon) > 0 \\ &\Leftrightarrow \quad a\varepsilon < b\varepsilon \\ &\Leftrightarrow^{b.A.} \quad a < b \end{aligned}$$

The last inequality holds by the initial assumptions, which completes the proof. ■

With this lemma in hand we can now proceed to derive benchmarks (upper and lower bound estimates) for the parameters of interest.

E.4.3 Upper Bound: Network Without Higher Order Spillovers.

In the "outbound" network higher order spills back to the originating nodes do not exist[91]: $HO_{\ell 0}$ and $HO_{\ell 1}$ would be 0. This is equivalent to assuming:

$$\mathbf{DiD} =^{\mathbf{b.A.}} \overline{\delta_1}\mathbf{D_t}(\mathbf{I} + \overline{\alpha}\mathbf{G} + \mathbf{0} + \mathbf{0} + ...) \tag{31}$$

which is equivalent to having[92]:

$$\begin{aligned} DiD_0 &= \overline{\delta_1} \qquad for\ treated\ L0 - nodes \\ DiD_2 &= 0 \qquad for\ L2 \\ & \qquad ...analogously\ for\ L3\ and\ higher \end{aligned} \tag{32}$$

By Lemma 1 this assumption leads to an upper bound of both coefficients. If all effects are of the same sign and $DiD_0 > DiD_1 > HO > 0$[93], the difference-in-differences for a node $\ell 1 \in L1$[94] would simply reduce to:

$$DID_1 \quad = \quad \frac{\overline{\alpha}}{NP_{\ell 1}}\overline{\delta_1} \tag{33}$$

A consistent estimator of $\overline{\delta_1}$ and the observed difference-in-differences will be enough to estimate $\overline{\alpha}$. In the "outbound network", I apply difference-in-differences on the level of directly treated nodes to obtain such an estimate. Then I move on to estimate $\overline{\alpha}$:

$$\begin{aligned} \hat{\overline{\delta_1}} &= \widehat{DiD_0} = \Delta\hat{\ell}0 - \Delta\hat{c}0 \\ \hat{\overline{\alpha}} &= \frac{\widehat{DiD_1}}{\widehat{DiD_0}}NP_{\ell 1} \end{aligned} \tag{34}$$

with $\Delta\hat{\ell}0 := \frac{1}{NP_{\ell 0}} * \sum_i(y_{i,\ell 0,t=0} - y_{i,\ell 0,t=1})$, $\Delta\hat{c}0 := \frac{1}{NP_{c0}} * \sum_i(y_{i,c0,t=0} - y_{i,c0,t=1})$. The definition of $\Delta\hat{\ell}1$ and $\Delta\hat{c}1$ for the $\widehat{DiD_1}$ parallels the definition of $\Delta\hat{\ell}0$ and $\Delta\hat{c}0$.

Discussion: The assumption in equation 31 implies no "multiplication-effects" or "feedback-loops" between the nodes.[95] In the light of the formalization presented here, this is a strong assumption. However, in the impact evaluation literature with fixed and stable classroom sizes or villages, this assumption is almost taken implicitely, whenever the researchers report merely

[91] Admittedly, in such a network, endogeneity would not be a problem in the first place.

[92] $D_{\ell 0}$ denotes the value of D at the central node, that is related to the focal node.

[93] DiD_0 (DiD_1) denotes the difference-in-differences for treated nodes (neighbors). For the reverse relationships ($DiD_0 < DiD_1 < HO < 0$) the estimate based on assuming an "outward bound" network gives a lower bound, if the effects go in opposite directions, my claims do not necessarily hold and will have to verified by the researcher. Slightly more involved assumptions will be needed.

[94] Which corresponds to an Indirect Treatment Effect or an "Externality"

[95] Neglecting higher-order spillovers is like implicitely introducing a temporal structure where a spillover takes time to occur and taking a snapshot after the first order effect. This is possible if, for example, spillovers are slow or if the temporal structure of the available data is fine grained enough.

the ATE and ITEs. (cf. Angelucci and De Giorgi (2009), Carmi et al. (2012), Dahl et al. (2012), etc. etc.).

Having said that, the upper bound estimator is quite suitable if higher order spillovers are negligible. In what follows I compute the lower bound estimates under the assumption of maximal higher order spillovers. This will give a sense of the maximal size of the bias that might result from assuming away the higher order complexities of a network.

E.4.4 Lower Bound: Network with Maximized Higher Order Spillovers.

In this subsection I derive the lower bound estimates under the assumption of a fully connected network. Formally, consider the matrix $\underline{\mathbf{G}}$, that corresponds to a fully connected network:

$$\underline{\mathbf{G}} = \begin{pmatrix} 0 & \frac{1}{N-1} & \frac{1}{N-1} & \cdots & \frac{1}{N-1} \\ \frac{1}{N-1} & 0 & \frac{1}{N-1} & \cdots & \frac{1}{N-1} \\ \frac{1}{N-1} & \frac{1}{N-1} & 0 & \cdots & \frac{1}{N-1} \\ . & . & . & & . \\ . & . & . & & . \\ \frac{1}{N-1} & \frac{1}{N-1} & \frac{1}{N-1} & \cdots & 0 \end{pmatrix}$$

First, observe that all nodes are direct neighbors, i.e. $NP_{\ell 0} = NP_{\ell 1} = NP_{\ell} = N-1$. Next, note that there are only two types of nodes: Directly treated nodes and neighbors. Let us now characterize the higher order spillovers that arrive at the treated node. From equation 25 we know that the spillovers that arrive at a node in L0 are given by:

$$\ell 0: \; DiD_0 = \delta_1(1 + 0 + \alpha^2 G_{ii}^2 + \alpha^3 G_{ii}^3 + ...)$$

The formula above points out that no spillovers of order 1 arrive at the treated node, since i does not link on to himself.[96] But in a network characterized by $\underline{\mathbf{G}}$, (and maintaining local treatment) the second order spillovers arrive from *every* neighbor, i.e. NP_ℓ times, third order spillovers arrive $(N-1)^2 - (N-1)$ times etc.[97] The number of channels for spillovers of order S is given by:

$$\begin{aligned} \#channels_{ii,S} &= (N-1)^{S-1} - (N-1)^{S-2} + (N-1)^{S-3} + ... \\ &= \sum_{s=1}^{S-1} (N-1)^s (-1)^{(S-1)-s} \qquad S \geq 2 \end{aligned}$$

The sum of second and higher order spillovers arriving at the treated node is:

$$\begin{aligned} HO_{ii} &= \sum_{S=2}^{\inf} \delta_1 \frac{\alpha^S}{(N-1)^S} \#channels_{ii,S} \\ &= \sum_{S=2}^{\inf} \delta_1 \frac{\alpha^S}{(N-1)^S} \sum_{s=1}^{S-1} (N-1)^s (-1)^{(S-1)-s} \end{aligned}$$

All non-treated neighbors are the same and the number of channels for spillovers of order S

[96] Note that this is precisely the point where the local treatment assumption is most useful, because had we treated T $>$ 1 nodes, then we would have to count T-1 direct spillovers that arrive at i, which obviously would render the following considerations less tractable.

[97] Counting the number of channels for third and higher order spillovers is a matter of combinatorics: The number of channels for higher order increases at an almost exponential rate, leading to potentially very large effects, that are moderated only by the decrease of the primary effects during transmission.

from node i to node j is computed almost[98] in the same way:

$$\begin{aligned}\#channels_{ij,S} &= (N-1)^{S-1} - (N-1)^{S-2} + (N-1)^{S-3} + ... \\ &= \sum_{s=0}^{S-1} (N-1)^s (-1)^{(S-1)-s} \qquad S \geq 2\end{aligned}$$

Again the sum of second and higher order spillovers at the neighboring nodes is:

$$\begin{aligned}HO_{ij} &= \sum_{S=2}^{\inf} \delta_1 \frac{\alpha^S}{(N-1)^S} \#channels_{ij,S} \\ &= \sum_{S=2}^{\inf} \delta_1 \frac{\alpha^S}{(N-1)^S} \sum_{s=0}^{S-1} (N-1)^s (-1)^{(S-1)-s}\end{aligned} \tag{35}$$

Before we can move on to derive the lower bound estimates, note that we have $\sum_{s=1}^{S-1}(N-1)^s(-1)^{(S-1)-s} < (N-1)^{S-1}$ which will be a convenient fact for simplifying the estimation of the lower bound.

$$\begin{aligned}HO_{ii} &= \sum_{S=2}^{\inf} \delta_1 \frac{\alpha^S}{(N-1)^S} \sum_{s=1}^{S-1} (N-1)^s (-1)^{(S-1)-s} < \\ &< \sum_{S=2}^{\inf} \frac{\alpha^S}{(N-1)^S} (N-1)^{S-1} = \\ &= \frac{1}{(N-1)} \sum_{S=2}^{\inf} \alpha^S = \frac{\alpha^2}{(N-1)} \frac{1}{1-\alpha}\end{aligned} \tag{36}$$

Let us call this expression $\overline{HO_{ii}}$. Analogously we obtain $\overline{HO_{ij}} = \frac{\alpha^2}{(N-1)} \frac{1}{1-\alpha}$. Plug these values into the equations 26 and 27 from above. With Lemma 1 at our disposal, we can use $\overline{HO_{ii}}$ and $\overline{HO_{ij}}$ to back out the lower bounds of the coefficients α and δ_1:

$$DiD_0 = \underline{\widehat{\delta_1}} + \overline{HO_{\ell 0}} \tag{37}$$

$$DiD_1 = \frac{\underline{\widehat{\alpha}}}{NP_{\ell 1}} \underline{\widehat{\delta_1}} + \overline{HO_{\ell 1}} \tag{38}$$

It is somewhat tedious, but straight forward to show, that solving this system of equations results in a quadratic equation for $\underline{\widehat{\alpha}}$:

$$\underline{\widehat{\alpha}}^2 - [\frac{DiD_0}{DiD_1} + (N-1)]\underline{\widehat{\alpha}} + (N-1) = 0 \tag{39}$$

The closed form solution for $\underline{\widehat{\alpha}}$ is hence given by:

$$\underline{\widehat{\alpha_{1/2}}} = \frac{1}{2}[\frac{DiD_0}{DiD_1} + (N-1)] + / - \sqrt{\frac{1}{4}[\frac{DiD_0}{DiD_1} + (N-1)]^2 - (N-1)} \tag{40}$$

Under weak regularity conditions[99] one solution is above 1 and another one between 0 and 1. The latter one is the solution for $\underline{\widehat{\alpha}}$ and it can easily be used to retrieve $\underline{\widehat{\delta_1}}$ from equation 26

Discussion: Note that this closed form solution requires only the number of nodes, and the two estimates from the difference-in-differences (for treated nodes and neighbors). It can be computed when nothing is known about the network, except how many agents and who was

[98] s now starts at 0.

[99] $DiD_0 > DiD_1$, which is to be expected for most treatments and follows from $\alpha < 0.5$ and $N > 1$

treated. It is thus as readily available as the upper bound estimators.

Clearly, one would immediately wish for more.[100] Having more information about the network structure or even the link strength between nodes is certainly desirable and, generally, will allow for more interesting additional results. Finally, while the proof here advantageously uses the local treatment assumption, I conjecture, that it is straightforward to extend it to treatments of more than one node.

F Aside: Reaction to Treatment of the Neighbor

Everything above was derived under the assumption that nodes do not observe or at least do not react to the local treatment of their neighbors. This is appropriate for neighbors of Wikipedia articles that get advertised on the start page.[101] In general however, subjects might observe treatment of their neighbors and react to the fact.

An example are children at school, who get annoyed or jealous when their peer was treated in a nice way and they were not.[102] In such situations the students/villagers might react to *merely observing* the treatment of their neighbors by selecting a different value for the outcome variable. To model such a situation we need to further augment the model in equation 14 by both the observable treatments (shocks) that are locally applied, and a term that captures the possible reaction to the treatment of the neighbor.

$$y_{it} = \alpha \frac{\sum_{j \in P_{it}} y_{jt}}{N_{P_{it}}} + X_{it}\beta + \gamma \frac{\sum_{j \in P_{it}} X_{jt}}{N_{P_{it}}} + \delta_1 D_{it} + \delta_2 \frac{\sum_{j \in P_{it}} D_{jt}}{N_{P_{it}}} + \epsilon_{it} \tag{41}$$

Where δ_1 measures the direct treatment effect and the new coefficient δ_2 measures reactions of the node, when it "observes" treatment of one (or several) of its peers. Consider again two connected nodes, where one is treated ($\ell 0$) in period t and the neighbors are not treated ($\ell 1 \in L1$). Assume for simplicity that $\ell 0$ is the only treated node in $\ell 1$'s neighborhood. Similarly, but different, we have:

$$\ell 0 :: y_{\ell 0t} = \alpha \frac{\sum_{j \in P_{\ell 0t}} y_{jt}}{N_{P_{\ell 0t}}} + X_{\ell 0t}\beta + \gamma \frac{\sum_{j \in P_{\ell 0t}} X_{jt}}{N_{P_{\ell 0t}}} + \delta_1 1 + \delta_2 \frac{\sum_{j \in P_{\ell 0t}} 0}{N_{P_{\ell 0t}}} + \epsilon_{\ell 0t} \tag{42}$$

$$\ell 1 \in L1 :: y_{\ell 1t} = \alpha \frac{y_{\ell 0t} + \sum_{j \in P_{\ell 1t}/\ell 0} y_{jt}}{N_{P_{\ell 1t}}} + X_{\ell 1t}\beta + \gamma \frac{\sum_{j \in P_{\ell 1t}} X_{jt}}{N_{P_{\ell 1t}}} + \delta_1 0 + \delta_2 \frac{1 + \sum_{j \in P_{\ell 1t}/\ell 0} D_{jt}}{N_{P_{\ell 1t}}} + \epsilon_{\ell 1t} \tag{43}$$

Now we get two types of spillover effects in this model: First the "pure spillover" α, due to the effect of treatment on the outcome of $\ell 0$. But second, also the "behavior change" of the node, δ_2, when it "observes" treatment of its peer kicks in.

Applying a Difference in Differences strategy alone will measure the joint effect of these two "spillovers". It will not identify α seperately, unless δ_2 is believed to be 0. If this assumption is not warranted only the total "treatment-of-peer"-effect can be measured. Depending on the application we might care about the effect of treatments, in which case this aggregate effect will be interesting. It is simply important to be aware that it is not possible to identify the pure spillover effect in such a setting.

[100] Note that if there is reason to believe that α is greater than 0.5 an analogue of Lemma 1 that relaxes my assumption of $\alpha < 0.5$ is required.

[101] For two reasons: (i) Wikipedia articles cannot react and (ii) the advertisement is not associated with any changes in the real world, so there is no reason for any updates.

[102] Other examples entail economic agents in a village, who observe that their neighbor was refused a social service for failure to comply with a requirement (e.g. sending their kids to school) or commuters in a city, who observe when their friends got caught (after the local transport authority increased the frequency of controls and the punishment for failure to present a valid ticket).